

Laurent Romary and Werner Wegstein

Consistent Modeling of Heterogeneous Lexical Structures

Warning

The contents of this site is subject to the French law on intellectual property and is the exclusive property of the publisher.

The works on this site can be accessed and reproduced on paper or digital media, provided that they are strictly used for personal, scientific or educational purposes excluding any commercial exploitation. Reproduction must necessarily mention the editor, the journal name, the author and the document reference.

Any other reproduction is strictly forbidden without permission of the publisher, except in cases provided by legislation in force in France.

revues.org

Revues.org is a platform for journals in the humanities and social sciences run by the CLEO, Centre for open electronic publishing (CNRS, EHESS, UP, UAPV).

Electronic reference

Laurent Romary and Werner Wegstein, « Consistent Modeling of Heterogeneous Lexical Structures », *Journal of the Text Encoding Initiative* [Online], Issue 3 | November 2012, Online since 15 October 2012, connection on 05 November 2012. URL : <http://jtei.revues.org/540> ; DOI : 10.4000/jtei.540

Publisher: Text Encoding Initiative Consortium
<http://jtei.revues.org>
<http://www.revues.org>

Document available online on:
<http://jtei.revues.org/540>

Document automatically generated on 05 November 2012.

TEI Consortium 2012 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

Laurent Romary and Werner Wegstein

Consistent Modeling of Heterogeneous Lexical Structures

1. Pooling Lexical Sources: A Digital Humanities Perspective

- 1 Our paper addresses the problem of interoperability between heterogeneous data sources, an issue that has regularly been the object of many debates within the Text Encoding Initiative (TEI) community and in general within many standardization groups providing models or formats for data interchange. At the core of the problem is the trade-off between expressivity—offering a flexible platform for representing a variety of possible structures—and processability—being able to predict under which conditions some data can be the object of a blind interchange, in particular in the context of them being processed randomly by a generic tool.
- 2 This trade-off has no generic solution, but it regularly arises in defining the components of such an expansive modeling platform as the TEI Guidelines. The TEI specifications are an expression of a balance of interests between the many, varied use cases from the community and the need to abstract away from such examples in order to design recommendations that new users can easily understand and apply in the context of their own encoding endeavours.
- 3 Throughout the TEI Guidelines one finds a stratification of corrections, constraints, and new features added over time, which have left some constructs as hybrid data models and which leave the user wondering which representation is the “optimal” one in a given context, leading to heterogeneous encoding practice in the global data space of existing TEI documents. Over the years, this has become more and more an issue as documents are increasingly accessible online and scholars increasingly collaborate on projects using TEI documents. That is, the “stratification” of the Guidelines has worsened the problem of interoperability.
- 4 In this paper we will focus on lexical structures, which we believe represent a typical case of the interoperability problem in terms of pooling data from heterogeneous sources. We have asked ourselves whether the TEI chapter dedicated to lexical data, simply entitled “Dictionaries,” should not be revised or at least be accompanied by further constraints on its usage so that basic operations related to the querying, displaying, or merging of lexical information could be made more straightforward.
- 5 From a digital humanities perspective, we want to understand if it is possible to find a balance between expressing precise constraints on the encoding of a primary source and leaving some freedom to the scholar who will see the encoding activity as a step in his research process. This is why we have made an attempt to identify a generic methodology for expressing encoding constraints on source texts based on the idea of local representation or *crystals* (Romary 2009). These crystals correspond to elementary constructs at a low level of granularity in a document, which, independently of the broader organization of the document itself, can be used to express a certain concept in an extremely regular way, thus making the further reuse of this information chunk easier. In this context, interoperability is related to the capacity of a person or a tool to process encoded crystals within a document independently of its origin.
- 6 After presenting the general background for modeling and representing lexical sources, we give an overview of the various crystals that form the basis of most existing types of lexical entries. For each of these crystals we make systematic recommendations with corresponding supporting arguments. In the second part of the paper we illustrate our proposals with concrete cases taken from various dictionary and lexical database projects.

2. Modeling Tools for Lexical Resources

- 7 The case of lexical data as presented in a dictionary offers an interesting experimental setting for studying interoperability in the context of standardisation. It is complex enough to reflect the variability which is intrinsic to the TEI Guidelines while providing a limited observational

setting for studying the granular structure of lexical entries as well as the rather high internal coherence that one specific lexical source usually has. Lexical resources also reflect the variety of analytical points of view that one may have on linguistic information, ranging from quite descriptive and verbose objects in the domain of standard human-oriented dictionaries to fully structured databases like those developed in the natural language processing domain.

In this paper we consider only lexical resources that are encoded semasiologically—where entries are determined according to the forms found in a language and further refined into the different senses that have been deemed relevant for this form. This *word-to-sense* organization is usually seen as the most appropriate for the representation of large coverage lexica, as opposed to onomasiological representations (*concept-to-term*), which better take into account the organization of domain-specific vocabularies (terminologies). The semasiological perspective is usually the underlying model for traditional print dictionaries as well as for large-scale lexica in the natural-language-processing domain (Halpern 2006; Atkins et al. 2002).

There are two main international standardization activities that are relevant for the modeling and the representation of semasiological resources: the Lexical Markup Framework (LMF) and TEI. In accordance with the modeling strategy of ISO committee TC 37, LMF (which has been standardised as ISO 24613:2008) provides a group of meta-models that can be combined to produce specific data models applicable to a wide range of lexical types or components including machine readable lexica, morphology, syntax, semantics, and multi-word expression. Even when the LMF specification provides a possible XML serialisation, it tends to be agnostic as to the actual implementation of the models it allows one to describe. On the other hand, the TEI has been seminal in offering a reference XML vocabulary for the representation of dictionaries, which is mostly compliant with LMF principles.¹ However, the variety of constructions that the TEI actually allows for the representation of the same lexical phenomenon could possibly be seen as a hindrance to the achievement of deep interoperability across heterogeneous lexical resources.

In this paper we take as a starting point the positions described by LMF and the latest release of the TEI Guidelines² in order to provide further insights into how to build lexical resources or dictionaries relying on a systematic use of standardised constructs. The work presented here is also based upon some core principles that have systematically guided our work, both theoretically but also practically, through the in-depth presentation of examples that have served as experimental background for testing our proposals. Even though the present work is not about modeling XML structures at large, several of these principles are derived from a more global concept of the kind of semantics that XML constructs convey and the way to actually reflect this in the design of XML formats.

With this perspective in mind, two generic constraints that affect the organization and semantics of lexical structures can be stated:

- Semantic grouping: Features that jointly convey a given meaning in a lexical entry should be systematically grouped together, even when only one such feature occurs and even at the cost of favoring more deeply-structured representations.
- Hierarchical dependency: Features, or groups thereof, which qualify a given level (for instance, an entry), are considered to be inherited by subcomponents (typically the senses) of the lexical entry unless otherwise stated (Ide, Kilgariff, and Romary 2000). (Here and below, we use “level” to refer to a hierarchical relationship within the data structure.)

From these constraints we will progressively derive specific recommendations for the local organization of lexical entries as guided by a crystal-based analysis. Comparing these with real data, and in particular with legacy dictionaries, we will try to understand possible transition schemes from weakly structured data to more standardized constructs.

3. Core Proposals: Towards a Systematic Description of Lexical Crystals

3.1. Crystals as Coherent Sub-structures

13 Introducing the concept of crystals in data modeling in general and in the TEI Guidelines in particular reflects the need to describe data structures that act as scaffolding for a coherent group of components (or elements in XML terminology). More precisely, a crystal can be defined as an *independent group of connected elements* (a *clique*) with semantic coherence. A typical example of a crystal is a structured bibliographical entry using the TEI's <biblStruct> element. This element contains internal structure (comprising <analytic>, <monogr> with <imprint>, and <series>), can be inserted at various places within the TEI architecture, and can be further expanded by other components or crystals (for example, <author>).

14 Without introducing any specific formalism here, we might define a crystal by:

- The set of mandatory and optional components that may occur in the crystal
- The structural organization of the crystal, stating in particular the hierarchical relations between components
- The anchor points of the crystal (<analytic>, <monogr> with <imprint>, and <series>), where it can be further expanded
- The global semantics of the crystal, in complement to the specific semantics of its component elements

15 A crystal is thus a modeling tool that can be used to provide a coherent description of a subset taken from a more complex data model (as is typically the case with the TEI Guidelines). To illustrate this, we will briefly demonstrate how the TEI Guidelines chapter on dictionaries can serve as a basis for implementing LMF, and point out some consequences this could have on the data architecture that we recommend for certain TEI elements.

16 As a starting point, let us consider the LMF subset depicted in figure 1, which implements the semasiological view of a lexical entry. This UML diagram states that a *Lexical Entry* is characterised by at least one *Form* component to which a hierarchically embedded series of *Sense* components may be associated. The *Form* component is further refined by means of an optional *Form Representation* component, which can be used to represent the various concrete implementations of a lexical form (e.g. phonetic, graphical, etc.). Finally, each component of the meta-model (corresponding here to a UML class) can be further characterised by properties attached to each of them.

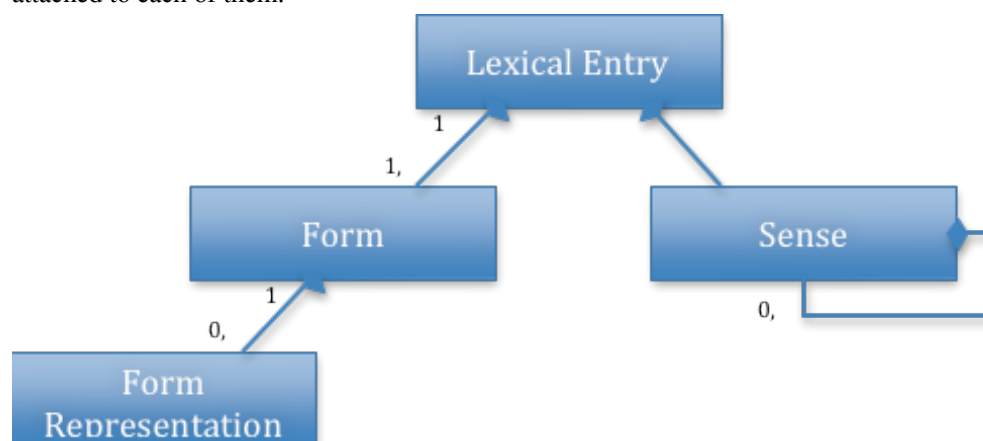


Figure 1: The Lexical Entry sub-structure of the LMF core package

17 Transposed to the TEI world, the LMF metamodel can be expressed as a TEI crystal rooted on the <entry> element. This crystal, depicted in figure 2, states that the minimal lexical entry in a sense as defined by TEI uses the <entry>, <form> and <sense> elements, with <form> being further decomposed by means of a series of elements implementing the Form Representation component of LMF.³ The picture also introduces three new classes, which could gather up all further descriptive elements needed to refine <entry>, <form>, and <sense>: model.entryDesc, model.formDesc, and model.senseDesc.

- 18 This first presentation of the TEI lexical entry as a crystal illustrates how this concept may help in describing complex structures that rely on constraints that go beyond (and deeper) than what we normally express by means of DTDs or schemas. Even though we do not systematically analyze the equivalences between LMF and the TEI in the following section, we hope that the preceding explanation will help the reader understand the logic behind the various constraints explained in subsequent sections. In a pattern analogous to the internal structure of the <cit> element, we see the organization of the various elements of this lexical entry crystal as a combination of a *structural description* (direct dependency of one element on another) and a *descriptive dimension* (further constraints applicable to the group of elements).

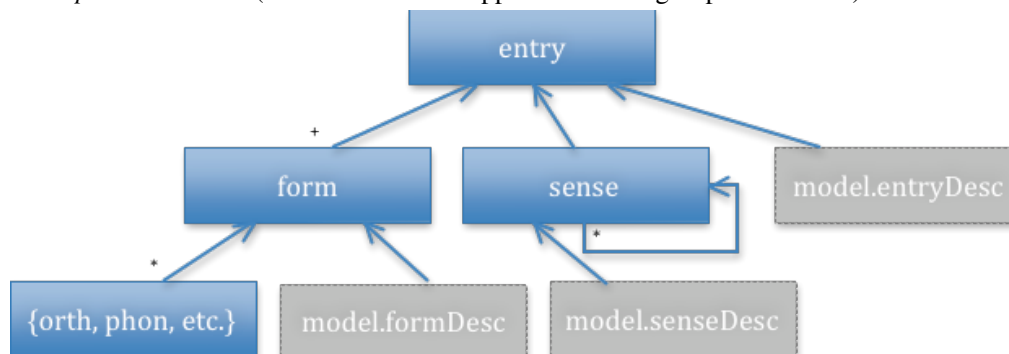


Figure 2: The ideal element-class organization of a TEI lexical entry

3.2. Morphographical Descriptions

- 19 In a semasiologically structured lexical entry, form information gives one or more realizations of a word—whether graphical, phonetical or iconical (by means of a picture or drawing)—which can be used to find the corresponding lexical unit. Such information may comprise abstract identifiers for the headword, namely the lemma, morphological components or categories (such as the consonantal pattern in Arabic), or any inflectional variant that can be associated with the entry.
- 20 The central issue in describing the corresponding morphographical crystal is that it should be based upon an abstract representation of *Form* as a component, which in turn groups together all the possible realizations of the corresponding form (the *Form Representation* component in LMF), as well as the associated constraints. In terms of good practices, one should thus refrain from providing a form representation (realization) in isolation and always include it within an embedding <form> element.⁴ Unless there is only one form associated with a given lexical entry, the form type (such as a lemma or inflected form) should be provided to ensure its univocal identification.
- 21 As a consequence, the minimal structure associated with a TEI-encoded lexical entry—where the only information given is that of a lemma (here, the French word *chat*; (en) *cat*)—should be encoded as follows:

```
<entry>
  <form type="lemma">
    <orth>chat</orth>
  </form>
</entry>
```

- 22 On this basis, additional variants of the form (such as pronunciation) can be added to the same form container, together with complementary information characterizing them. For instance, when more than one orthography is used to provide the form, the appropriate @type attribute should be used to qualify the corresponding orthography. In the following example, the lemma for the Korean word “##” (chida; (en) *to hit*) is provided in Hangul orthography ((ko) ##) orthography together with a Romanized form.

```
<form type="lemma">
  <orth type="##">##</orth>
  <orth type="romanized">chida</orth>
</form>
```

- 23 As a next step, we advocate the definition of stable values for the @type attribute on <orth>, adopting ISO 15924 to refer to the script.
- 24 When alternative forms are provided, indicating, for example, inflectional variation, then the variants should be encoded in full in order to reflect linguistic differences. For instance, the example provided in Annex B of LMF (*clergyman*) is reformulated in TEI as follows:

```
<entry>
  <gramGrp>
    <pos>commonNoun</pos>
  </gramGrp>
  <form type="lemma">
    <orth>clergyman</orth>
  </form>
  <form type="inflected">
    <orth>clergyman</orth>
    <gramGrp>
      <number>singular</number>
    </gramGrp>
  </form>
  <form type="inflected">
    <orth>clergymen</orth>
    <gramGrp>
      <number>plural</number>
    </gramGrp>
  </form>
</entry>
```

3.3. Grammatical Information

- 25 Grammatical information may appear at various points within a dictionary entry; it is there to provide additional information about the core objects comprising the entry. In the lexicographic tradition grammatical information qualifies the lemma, or rather, since the lemma is just a code representing the entry as a whole, syncretizes the grammatical features that apply by default to all possible occurrences of the word. However, the grammatical information can also occur at many other possible levels of the entry, qualifying inflected forms in a more precise way (as in the “clergyman” example above), indicating specific constraints associated to a sense, or even qualifying the occurrence within an example of phrasal expression. As a whole, a grammatical crystal defined according to these principles may be used at any place where the usage of a word is described.
- 26 The notation for grammatical features within human-oriented dictionaries varies greatly: a given grammatical constraint can, for instance, be represented by a prototypical morpheme (e.g. *der / die / das* to indicate grammatical gender in German) or by means of a descriptive phrase (*used in the plural form*). At best, idiosyncratic codes are used (e.g. *masc.*, *fém.*), though they are not always consistently applied within a single dictionary, let alone across dictionaries. There is no doubt that such a situation prevents one from querying lexical entries that include grammatical constraints in a coherent way. It is therefore a priority to establish requirements for the representation of grammatical features in a way that is both standard and yet preserves the initial editorial choices. As a basis for such recommendations we recommend that TEI-based encoding of dictionary entries should be in keeping with the following elementary principles:
- Grammatical features should systematically be embedded within a <gramGrp> container element, even if only one feature is present and even if the grammatical information is split up so that more than one <gramGrp> container may be necessary.
 - Whereas one should be flexible with the textual content of a grammatical descriptor, it is of utmost importance to normalize the intended value by means of a @norm attribute.
- 27 For instance, when a value for the grammatical gender is given by means of a determiner, the @norm attribute will provide the reference value (e.g. as a code from the ISOcat data category registry).⁵ Depending on the encoder’s editorial choices, a minimal encoding might look like the following example:

```
<form type="lemma">
```

```

    <gramGrp>
      <gen norm="feminine">die</gen>
    </gramGrp>
    <orth>Katze</orth>
  </form>

```

28 A more elaborate encoding scheme could lead to the following lemma structure:

```

<form type="lemma">
  <form type="marker">
    <gramGrp>
      <pos norm="determiner"/>
      <gen norm="feminine"/>
    </gramGrp>
    <orth>die</orth>
  </form>
  <form type="head">
    <gramGrp>
      <pos norm="noun"/>
      <gen norm="feminine"/>
    </gramGrp>
    <orth>Katze</orth>
  </form>
</form>

```

29 In general, such grammatical descriptions should be thought of as being equivalent to the provision of feature structures and thus mappable onto an <fs> element. For instance, the preceding minimal encoding example (omitting the orthographic form) is equivalent to:

```

<fs>
  <f name="gender"><symbol value="feminine"/></gen>
</fs>

```

30 The next stage in providing a recommendation is to make sure that values for the @norm attribute are stable within a project and, when possible, across projects. We recommend two complementary strategies:

- For a given project, document and publicize the values used for the norm attribute so that the community may be aware of possible discrepancies
- Relate such values to entries in the ISOcat data category registry so that they are mapped onto standardized conceptual references.

31 It should be noted that at the time of writing, there is an item on the TEI Council agenda to better integrate mechanisms available in ISO 12620:2009 (the standard which defines the structure of ISOcat) within the TEI architecture to facilitate such mappings. We can thus expect that these recommendations may become in due course standard practice within the TEI community.

3.4. Senses as Systematic Entry Points

32 The representation level introduced by the Sense component in LMF and its counterpart <sense> in the TEI Guidelines is an essential concept implementing the semasiological perspective of a dictionary. Still, a “lazy” encoding style for dictionary entries could lead to the idea that such a structure is superfluous when, for instance, a word can directly be described at the same level as the morphological and grammatical information by a simple definition or a translation that is a child of <entry>. Indeed, it is often the case in the simplest forms of legacy lexical structures that senses are not explicitly separated out in the microstructure of the entry. We consider this bad practice and recommend that <sense> be used to enclose *all* descriptors that describe the *signified* (as opposed to the signifier, that is the <form>, in the Saussurian sense).

33 As can be observed from the variety of constraints that may apply to a <sense> element within a lexical entry, the underlying understanding of the semasiological model extends to the organization of senses that do not rely on strict semantic criteria (Ide, Kilgariff, and Romary 2000). This is not so much of a paradox when we think of the numerous ways by which semantic variation may be observed, among which we can include pure morpho-syntactic or syntactic markers. As a result, we consider that <sense> should be used to describe any

subdivision reflecting a variation in usage for a given word. In an extreme case, applying automatic collocation extraction tools (Kilgarriff and Tugwell 2002) may result in generating lexical entries automatically where senses correspond to the various collocation classes that the tool has determined.

34 We thus see the sense component in LMF and the <sense> element in TEI as a generic container organizing the further description of a signifier, which may contain information related to:

- The actual syntactico-semantic restriction applicable to the sense being described, for instance by means of further grammatical constraints, a definition, or some usage restriction
- The provision of further illustrative information, in particular contextualized examples or translations (see the section on the <cit> element below)
- Relational information referring to external information expressing the same meaning, either within another lexical entry or an external ontological reference (such as in the lexical database project WordNet, described by Miller and Fellbaum [2007]).

35 In order to actually facilitate further querying, it is important that each feature intended to be associated with a sense shall be precisely typed. Precise typing requires that clearly defined typologies be associated with elements such as <usg> and <cit>. Furthermore, dictionary projects should be able to document precisely how much restrictive or illustrative information is inherited along embedded senses. For instance, a clear editorial strategy should state whether grammatical constraints replace or complete existing ones at a higher level of a sense hierarchy.

3.5. <cit>: A Generic Linguistic Quotation Tool

36 The <cit> element in TEI P5 is the result of a merger of several constructs from former editions of the TEI chapter on dictionaries that had been created to handle examples and translations in dictionary entries. The underlying aim of the new framework was twofold. On the one hand, the objective was to provide greater coherence to the way language excerpts appear not only in dictionaries but in textual content in general. On the other hand, the TEI Council wanted to design a sound framework for dealing with additional references or constraints provided in a lexical entry to compliment the quoted object itself, taking into account that such refinements may lead to recursive constructs. In terms of interoperability across TEI-based applications, the main vision behind the <cit> element, and the crystal it shapes, is to provide entry points for generic searches for quoted language in texts, from the point of view both of the full-text content and of providing a systematized representation of constraints associated with the full text.

37 Language quotations in text may indeed take many different forms. In dictionaries the most basic quotation is simply a phrase or sentence exemplifying the headword. Most of the times, this quotation does not appear alone but is refined according to two main axes:

- Indication of the source of the quotation, for instance the following from P5 2.0.0: *‘La valeur n’attend pas le nombre des années’* (Corneille)
- Provision of usage information, stating constraints that the example is bound by, such as domain or pronunciation, as in the following from P5 2.0.0: **some** ... 4. (*S~* and *any* are used with *more*): Give me ~ more/s@'mO:(r)/

38 In the case of multilingual dictionaries, language quotations are similarly used to provide equivalences for the entry (or sub-sense thereof) in the target language. In a way that is similar to the monolingual case, further refinement of the encoding structure of a quotation may indicate some source or usage information, but it may also document the target language proper. A usual case here is the indication of the grammatical gender of a noun equivalent in the target language.

39 Quotation constructs are not covered in LMF but can easily be modeled as an extension to the LMF core packages. Figure 3 is a simple representation for such an extension. The approach is similar to the one we advocate above for grammatical information in relation to senses, in which the quoted text is embedded in a quotation construct even if no refinement is actually stated.

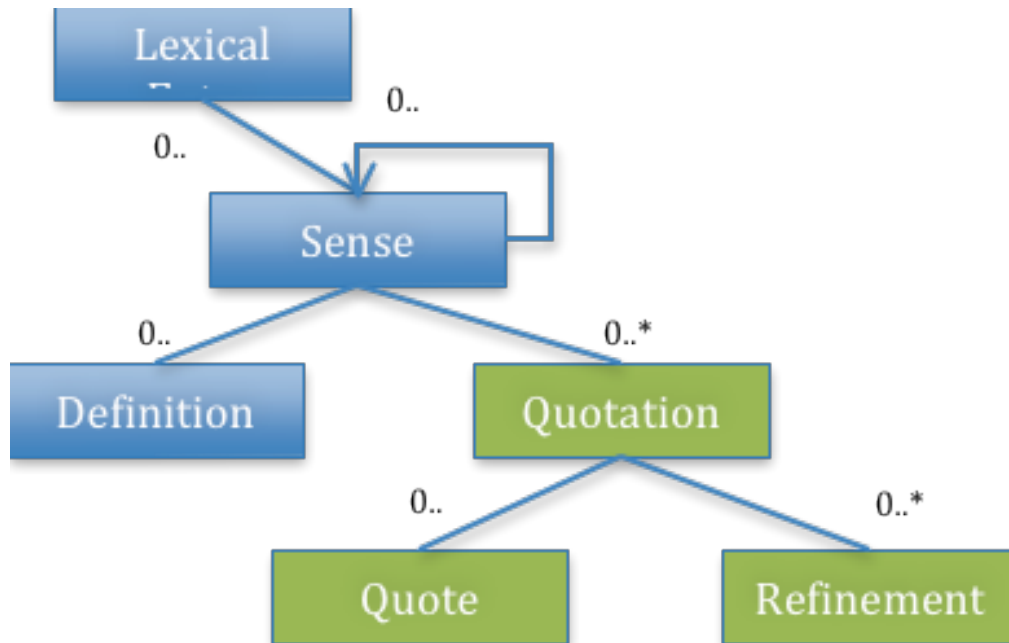


Figure 3: An LMF extension for quotations represented in a dictionary

In the TEI Guidelines, the quotation construct is implemented by means of the <cit> element, which has the following characteristics:

- The quoted object may be realized not only by means of a <quote> or <q> (both from the model.qLike class) but also as a more elaborated construct such as an XML object (<egXML>, a member of model.egLike).
- The refinement of a quotation can be instantiated as a bibliographic reference (using an element from model.biblLike), as a pointer or external reference to a constraint (using an element from model.ptrLike), as specific lexicographic features such as grammatical constraints (using an element from model.entryPart), or through the inclusion of feature structures in <cit>—accidental by design—which are part of model.global. It should be noted that a refinement can actually be an embedded <cit> (by virtue of the inclusion of model.entryPart in the content model of <cit>), thus offering, for example, a natural way to provide a translation of a quotation.

Note that the TEI Guidelines already systematize the values of the @type attribute to “example” and “translation” for use in dictionaries.

Given the variety of possible cases where <cit> may be used and the potentially infinite combinations of refinement, it may be difficult to provide clear requirements for its application. Basically a proper usage of <cit> should allow a human reader or a processor to identify one quoted object and treat all other components as refinements in which semantics are understood in a conjunctive way (in other words, all refinements apply *en bloc* to the quoted object). By default, the quoted object should be the first child of the <cit> element or, in general, the first child that is a member of either model.qLike or model.egLike.

Although the second part of this paper provides several applications of <cit> in the context of our observational corpus, we can illustrate here some basic usages of this element from examples available in the TEI Guidelines.

In the following prototypical case, a simple example for the headword is associated with a refinement giving the pronunciation of part of the quoted text:

```

<cit type="example">
  <quote>Give me <oRef/> more</quote>
  <pron extent="part">s@'mO:(r)</pron>
</cit>

```

The next example illustrates the representation of a translation refined with a grammatical feature:

```

<cit type="translation" xml:lang="fr">
  <quote>habilleur</quote>

```

```

    <gramGrp>
      <gen>m</gen>
    </gramGrp>
  </cit>

```

- 46 Finally, we cannot resist presenting a recursive case where the embedded <cit> is used as an additional descriptive element for the quoted text at the higher level:

```

<cit type="example">
  <quote>she was horrified at the expense.</quote>
  <cit type="translation" xml:lang="fr">
    <quote>elle était horrifiée par la dépense.</quote>
  </cit>
</cit>

```

4. Illustrated Guidelines for Early Printed Dictionaries

4.1. Lexicographical Justification

- 47 We tested our encoding concepts using printed dictionaries from the second half of the 18th century for two reasons. First, in the history of English lexicography the early 18th century marks the beginning of modern dictionary practice (Landau 2001, 60–66). Samuel Johnson’s *Dictionary of the English Language*, first published in 1755, perfectly embodies these advances in lexicography. Johnson is the first English lexicographer to include thousands of other quoted “‘authorities’ within his text as illustrations of word use” (Reddick 1996, 9). His dictionary also brought together “for the first time key conventions for future dictionary presentation: the folio⁶ design is a system of typography that displays the structure of each entry, though there are inconsistencies of abbreviation and ambiguities” (Luna 2005, 193). Thus this dictionary offers an ideal test bed to study problems in providing a consistent encoding in P5 of a source document that offers notational inconsistencies. Second, because Johann Christoph Adelung⁷ translated Samuel Johnson’s dictionary into German (Adelung 1783–1796), Johnson’s dictionary opens up additional perspectives for the study of bilingual lexicographical resources in the 18th century and research into the history of revision and the reuse of dictionaries.
- 48 We test our modeling of lexicographic structures with three samples from Johnson’s monolingual dictionary representing the most frequent word-classes: the adjective ABLE, the verb To APPLAUD, and all entries for the noun APPLE (the use of all caps versus small caps by Johnson is explained below). We further compare Johnson’s *apple* entries with the section of *apple* entries in Adelung’s bilingual English-German translation of Johnson’s dictionary. To illustrate the differing encoding structures of bilingual German-English dictionaries we use Eber’s entry FÄHIG, the equivalent of ABLE. As a source for this entry, Ebers obviously used only the German-French dictionary of Christian Friedrich Schwan (Schwan 1782), so we include Schwan’s entry FÆHIG in order to illustrate dictionary reuse across languages in the 18th century. The images of the encoded pages are given as a supplement to this article.

4.2. Typographic Analysis and Text Encoding

- 49 Luna begins his essay on the typographic design of Johnson’s dictionary with some reflexions on how a typographer would analyze a dictionary: “In particular, how does a typographer look at a dictionary that is also a cultural artifact, as Samuel Johnson’s *Dictionary of the English Language* undoubtedly is?” (2005, 175). Building on a more wide-ranging definition of typography as “configuration of verbal graphic language,” Luna concludes, “the main concern of this essay is not the quality of the printing, nor the nature of the paper, nor even the origin of the founts of type used to compose the *Dictionary*, but how its visual presentation reflects the structure of the text, its usability, and perhaps even its compiler’s intentions” (2005, 175).
- 50 This concept comes very close to what a TEI encoding of a dictionary in an adequate granularity should achieve: reflecting the structure of the encoded text, facilitating re-usability in electronic form and—at its best—assisting in the detection of the author’s intentions. In

order to put our aim of a consistent modeling of heterogeneous structures into practice, we follow some basic principles.

We adopt a conservative editorial view for our literal transcription (see section 9.5.1 of P5) and try to keep the latter close to the printed original: we do not add any character to the original text or delete it, we transcribe the text in the order in which it appears in the source, we preserve the linear structures of the text with <pb>, <cb> and <lb>, and we retain the end-of-line hyphenation (see section 3.2.2 of P5). With such orthographical variation within the texts of the dictionaries, this makes transcription much easier. For clarity and to ensure a consistent encoding we encode only a few structurally important typographic features (significant use of typeface and italics) at the level of the lexical entry.⁸

4.3. Encoding Practice at the <entry> Level

With re-usability, interoperability, and sustainability of the dictionary entries in mind, we use two attributes to refine the <entry> element: @xml:id to guarantee a robust and reliable non-ambiguous identification and @type for classification of the entries.

The @xml:id attribute is composed of four parts, each separated by a dot:

1. two initials of the author's name and a combination of six letters or numbers to identify the encoded edition precisely
2. four digits for the year of publication
3. six digits for the running number of the entry (given as a random value in the examples)
4. the lemma, transcribed in lower case only and with any incidental spaces replaced by underlines.

Thus our sample entry ABLE in Samuel Johnson's dictionary is assigned the @xml:id 'sjdict1f.1755.000123.able'. In the first part, "sj" is taken from Samuel Johnson, "dict" reflects the title *Dictionary of the English Language*, and "1" indicates the edition and "f" the format folio (because edition and format are both rather important for a precise identification of the different printed editions of Johnson's dictionary). They are not necessary for Adelung (Henne 2001, 170), Ebers (Lewis 2012), and Schwan.

We use the TEI @type attribute of <entry> to distinguish typographically or verbally marked types of entries and map them onto corresponding identifiers of the ISOcat data category registry. The @type attribute used on <entry> belongs to the attribute class att.entryLike, which includes a list of suggested values for @type. For the entries in Johnson's *Dictionary* we had to add some more fine-grained distinctions to the list of suggested values.

An occasional user of Johnson's *Dictionary* may be puzzled about the typesetting of entry headwords. Thus APPLAUD and APPLE are in full caps, while APPLAUSE and APPLE TREE are in small caps. Now and then, however, entries appear typeset in italic capital letters, e.g. *ABORIGINES* and *ABRACADABRA*. In his preface, Johnson explains the background for these marked differences, which for him reflect basic lexicographical distinctions: "In the investigation both of the orthography and signification of words, their ETYMOLOGY was necessarily to be considered, and they were therefore to be divided into primitives and derivatives. A primitive word, is that which can be traced no further to any *English* root; . . . Derivatives, are all those that can be referred to any word in English of greater simplicity" (1755, 3f). Thus primitives or roots are marked by full caps and the derivatives by small caps. Furthermore, the entries in italic capital letters indicate foreign words used in the English language (Luna 2005, 181).

As Luna notices (2005, 196 fn. 24), this distinction of entries echoes a completely different way of organizing a dictionary: word-families, represented by roots (in alphabetical order), followed by their derivatives (ordered non-alphabetically into morphological or etymological groups). Since Johnson used a single alphabetical order for all entries, this organizing principle is no longer clearly visible. It is only faintly reflected in the differentiation of the lemmas. But it is still implicit and that is why we think it should be encoded explicitly as a significant feature of the dictionary structure. Accordingly, we map the entries representing lexical units in Johnson's *Dictionary* onto the ISOcat identifiers *root* or *derivation* and use *foreign* to indicate foreign words respectively. Two examples: ABLE and APPLE *of Love*.

```

<entry xml: type="Root">
  <form type="lemma" norm="able">
    <lb/><orth rend="allcaps">A'BLE</orth><pc>.</pc>
    <gramGrp><pos norm="adjective">adj.</pos></gramGrp>
  </form>
<entry xml: type="Phrase">
  <form type="lemma" norm="apple of love">
    <lb/><orth><hi rend="smallcaps">APPLE</hi> <hi rend="italics">of
    Love</hi></orth><pc>.</pc>
    <gramGrp><pos norm="noun"/></gramGrp>
  </form>
  <sense>
    <cit type="Encyclopedic_Information">
      <quote><lb/>Apples of love are of three sorts; ...
      <bibl><author>Mortimer</author>'s <title>Husbandry</title>.</bibl>
    </cit>
  </sense>
</entry>

```

- 58 The typography of the entry APPLE *of Love*—small caps for *apple* though belonging to the root entries, italics for *of love*, and the word class information missing from the source (though supplied in the encoding)—indicates uncertainty about the word status of the entry. Furthermore, the classification as type *phrase* may require some explanation. Valerie Adams comments in her introduction to word-formation on the distinction between words and phrases: “Certain noun-preposition-noun phrases also show their incomplete unification by the possibility of pluralizing the first noun” (1976, 9). Since the illustrative quotation of Mortimer’s book on Husbandry starts with the plural form “apples”, we regard the type “Phrase” here as justified and did not consider alternative ISOcat options.

4.4. The <form> Block

- 59 The <form> element is designed to contain information on the written form (encoded using <orth>) and, if present, the spoken form (encoded using <pron>) of one lemma. We use <form> with two attributes: a @type attribute to distinguish the lemma from any given inflected forms and a @norm attribute to even out any orthographic variation, such as the use of upper or lower case, hyphenation, or special markers to indicate the stress position within the orthographic representation of the lemma. The <form> block contains a number of elements including <orth> and <gramGrp>; the TEI <stress> element, designed for stress patterns given separately, is not applicable here, apart from the fact that we did not want to split up the orthographic representation any further or change it.
- 60 Within <orth>, typographic details are stored in a @rend attribute. In Johnson’s *Dictionary* we use it to store his typographic differentiation of the printed entries: that is, his distinction between all caps and small caps. In Schwan’s dictionary it is used to distinguish two different orthographic representations of the German lemma, the first with Antiqua capital letters only, the second with upper and lower case, depending on the German orthography, and using a Fraktur typeface.
- 61 We use <gramGrp> to collect grammatical information such as part-of-speech (in a <pos> element) or gender (in a <gen> element). Quite often, grammatical information precedes or follows the orthographic representation of the entry, such as the infinitive marker *To* in entries for verbs in Johnson’s dictionary or the determiner *der*, *die*, *das* in German noun entries. We capture this information with a <gram> element and a @type attribute containing the appropriate ISOcat value. Without exception, we store all elements that interpret grammatical features like <pos>, <gen>, or <gram> within a <gramGrp> element, once again using a @norm attribute to map the different grammatical descriptions given in the dictionaries to an ISOcat entry. This way, we avoid conflicts with the order of text on the printed page and can adjust inconsistencies like missing word class information, such as by adding an empty <pos> element with a @norm attribute based on information collected elsewhere in the entry. One example is Johnson’s entry APPLAUD that requires two <gramGrp> elements to capture the grammatical information:

```

<p> n="148"/><cb n="APP"/>

```

```

<entry xml: type="Root">
  <lb/><form type="lemma" norm="applaud">
    <gramGrp><gram type="infinitiveParticle">To</gram></gramGrp>
    <orth rend="allcaps">APPLA'UD</orth><pc>.</pc>
    <gramGrp><pos norm="verb">v.a.</pos></gramGrp>
  </form>
  <etym>
    <pc>[</pc><mentioned xml:lang="la">applaudo</mentioned><pc>,</pc>
      <lang><abbr>Lat.</abbr></lang><pc>]</pc>
  </etym>
  <lb/><sense>
    <num>1.</num>
    <def>To praise by clapping the hand.</def>
  </sense>
  <lb/><sense>
    <num>2.</num>
    <def>To praise in general.</def>
  </sense>
  <cit type="example">
    <lb/><quote>I would applaud thee to the very echo,
    <lb/>That should applaud again.</quote>
    <bibl><author><abbr>Shakesp.</abbr></author><title>Macbeth</title>.</bibl>
  </cit>
  <cit type="example">
    <lb/><quote>Nations unborn your mighty names shall sound,
    <lb/>And worlds applaud that must not yet be found!</quote>
    <bibl><author>Pope</author>.</bibl>
  </cit>
</entry>

```

- 62 Our use of <pc> is governed by the principle that we avoid punctuation marks as delimiters of
 text in elements within <form> and within <etym>; this is for ease of reusability and searching.
- 63 In testing our encoding concept we encountered some phenomena—word class in grammar
 and hyphenation in orthography—which prompted us to reinforce our aim of consistently
 modeling heterogeneous lexicographical data through normalization. The first case has to do
 with an old problem of word classes: the categories of adjective and adverb in German. Ebers
 defines the part-of-speech information in his entry *fähig* with the abridged terms in Latin *adj.*
et adv. This concept—one word, two word classes—is not compatible with the present-day
 understanding of word classes in German: since adverbs in German are never inflected and
fähig is capable of inflection, this word is generally regarded as an adjective in any present-
 day dictionary of German. Of course, we do not alter Ebers' word class definition, but we
 suggest resolving the word class conflict in this and in comparable cases by standardizing the
 value of the @norm attribute on <pos>, using the ISOcat value *adjective* only. Ebers' example
 entry *fähig* in abridged form:

```

<entry xml: type="main">
  <form xml:lang="de" type="lemma" norm="fähig">
    <lb/><orth>Fähig</orth><pc>,</pc>
    <gramGrp>
      <pos norm="adjective" xml:lang="la">adj. et adv.</pos>
    </gramGrp>
  </form>
  <sense> ... </sense>
</entry>

```

- 64 The second phenomenon has to do with hyphenation, an old problem primarily but not only in
 the English language. First, consider Johnson's noun compounds with *apple* in abridged form:

```

<entry xml: type="derivation">
  <form type="lemma" norm="apple graft">
    <lb/><orth rend="smallcaps">APPLE-GRAFT</orth><pc>.</pc>
    <gramGrp><pos norm="noun">n.s.</pos></gramGrp>
  </form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.009999.graft">graft</mentioned>
  </etym>

```

```

    <pc>.]</pc>
  </etym>
<sense> ... </sense>
</entry>
<entry xml: type="derivation">
  <form type="lemma" norm="apple tart">
    <lb/><orth rend="smallcaps">APPLE-TART</orth><pc>.</pc>
    <gramGrp><pos norm="noun"/></gramGrp>
  </form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.029999.tart">tart</mentioned>
    <pc>.]</pc>
  </etym>
<sense> ... </sense>
</entry>
<entry xml: type="derivation">
  <form type="lemma" norm="apple tree">
    <lb/><orth rend="smallcaps">APPLE TREE</orth><pc>.</pc>
    <gramGrp><pos norm="noun"><abbr>n.s.</abbr></pos></gramGrp>
  </form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.039999.tree">tree</mentioned>
    <pc>.]</pc>
  </etym>
<sense> ... </sense>
</entry>
<entry xml: type="derivation">
  <form type="lemma" norm="apple woman">
    <lb/><orth rend="smallcaps">APPLE WOMAN</orth><pc>.</pc>
    <gramGrp><pos norm="noun"><abbr>n.s.</abbr></pos></gramGrp>
  </form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.049999.woman">woman</mentioned>
    <pc>.]</pc>
  </etym>
  <sense> ... </sense>
</entry>

```

65 Apart from the special case “APPLE *of love*,” both “APPLE-GRAFT” and “APPLE-TART” are hyphenated, whereas “APPLE TREE” and “APPLE WOMAN” are spelled as two separate words. There is no consistent distinction here between open (word-spaced) and hyphenated compounds. Noel Osselton gives a compact *résumé* of “variation of hyphenated compounds” in entries and their steady downgrading in the second half of the dictionary from the letter M onwards (2005). Against this background we have used the @norm attribute of <form> in order to provide the best support for search procedures: we have retained the original hyphenated and open compound spellings from Johnson’s text but have encoded the open or word-spaced form on the @norm attribute as the standardized form.

66 In his translation of Johnson’s apple entries, Adelung takes a different view. He unifies the hyphenated spelling for all the apple compounds, downgrades the hybrid entry *Apple of love* to appear as a form mentioned within the base entry apple, and adds more compounds, taken from other sources mentioned in the introduction:

```

<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple">
    <lb/><orth>'Apple</orth><pc>,</pc>
    <gramGrp>
      <pos norm="noun" xml:lang="la">subst.</pos>
    </gramGrp>
    <pc>(</pc><pron>äpp'l</pron><pc>,</pc>
  </form>
  <etym><mentioned><lang xml:lang="ang">angels.</lang>

```

```

    <lang xml:lang="nds">niederd.</lang>aep- <lb/>pel</mentioned>
    <pc>,</pc> <mentioned><lang xml:lang="de">deutsch</lang> Apfel</mentioned>
    <pc>.</pc><pc></pc></pc>
  </etym>
  <sense xml:lang="de">
    <num>1</num>
    <def>Die Frucht des <lb/>Apfelbaumes,</def>
    <cit type="translation"><quote>der Apfel.</quote></cit>
  </sense>
  <sense xml:lang="de">
    <num>2</num>
    <cit type="Encyclopedic_Information">
      <quote>Wegen eini-<lb/>ger Ähnlichkeit in der Gestalt ...</quote>
    </cit>
    <cit type="Encyclopedic_Information">
      <quote><mentioned xml:lang="en">The Apple of love, Love-apple</mentioned>
        o-<lb/>der <mentioned xml:lang="en">Wolf's Peach</mentioned>,</cit>
      <cit type="translation" xml:lang="de"><quote>Liebesapfel</quote>
    </cit>
    <term xml:lang="la">Lycoper-<lb/>sicon<name nymRef="Linné">Linn.</name>
    </term>auch wohl eine Art des <term xml:lang="la">Sola-<lb/>num</term>;
    <mentioned xml:lang="en">the Mad-apple</mentioned>,</sense>
  <sense xml:lang="de">
    <num>3</num>
    <usg>Figürlich,</usg><def>die Pupille in dem Auge,</def>
    <cit type="translation"><quote>der <lb/>Augapfel,</quote></cit>
    <xr type="synonym "><lbl>welcher wohl auch
      <ref xml:lang="en" target="#adwbengl.1783.009999.eye-ball">
        Eye-ball</ref> ge-<lb/>nannt wird.</lbl>
    </xr>
  </sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple coar">
    <lb/><orth>'Apple-coar</orth><pc>,</pc>
    <gramGrp<pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
  </form>
  <etym><lbl>von</lbl>
    <mentioned xml:lang="en" corresp="#jagkwbed.1783.000999.apple">
      apple 1)</mentioned>
  </etym>
  <sense>
    <def>der Griebs oder Gröbs in dem Apfel.</def>
  </sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple graft">
    <lb/><orth>'Apple-graft</orth><pc>,</pc>
    <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
  </form>
  <sense>...</sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple loft">
    <lb/><orth>'Apple-loft</orth><pc>,</pc>
    <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
  </form>
  <sense>...</sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple monger">
    <lb/><orth>'Apple-monger</orth><pc>,</pc>
    <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
  </form>
  <sense>...</sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple paring">
    <lb/><orth>'Apple-paring</orth><pc>,</pc>

```

```

        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple roaster">
        <lb/><orth>'Apple-roaster</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple squire">
        <lb/><orth>'Apple-squire</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple tart">
        <lb/><orth>'Apple-tart</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple thorn">
        <lb/><orth>'Apple-thorn</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple tree">
        <lb/><orth>'Apple-tree</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>
<entry xml: type="main">
    <form xml:lang="en" type="lemma" norm="apple woman">
        <lb/><orth>'Apple-woman</orth><pc>,</pc>
        <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
    </form>
    <sense>...</sense>
</entry>

```

- 67 These examples illustrate that, despite differences in detail, the <entry> and <form> information can be encoded using the same pattern. Missing standard information (like word class) can be supplied without modification of the transcription of the printed text. Even if the encoding cuts into typographical structures (such as <pron> in Adelung's dictionary), it does not corrupt the transcription.

4.5. <etym>: Between Etymology and Word-Formation

- 68 As noted above, Johnson emphasized the importance of etymology in his preface. Accordingly, he opens his dictionary with a grammar, and, in the introduction to the chapter "Of DERIVATION", explains: "That the English language may be more easily made understood, it is necessary to enquire how its derivative words are deduced from their primitives, and how the primitives are borrowed from other languages" (1755, 47). In compound word entries, he uses square brackets following the part-of-speech information to mark the root components of the compound—his derivatives (for example, in APPLE-GRAFT: [from *apple* and *graft*]); for root entries, he provides information about related words in Indo-European, Romance or Germanic languages, if necessary with an English translation (for example, in ABLE: [*habile*, Fr. *habilis*, Lat. Skilful, ready.]). In accordance with Johnson's method, we use the <etym> element for both cases. The <etym> element requires no additional attribute to distinguish these two cases

since its content structure clearly indicates to what type of entry a given <etym> belongs and how it is to be interpreted:

```
<entry xml: type="Root">
  <form>...</form>
  <etym><pc>[</pc>
    <mentioned xml:lang="fr" >habile</mentioned><pc>,</pc>
      <lang><abbr>Fr.</abbr> </lang>
    <mentioned xml:lang="la">habilis</mentioned><pc>,</pc>
      <lang><abbr>Lat.</abbr></lang>
    <lb/><gloss xml:lang="en">Skilful<pc>,</pc> ready<pc>.</pc>
    </gloss><pc>]</pc>
  </etym>
```

- 69 In the encoding of the entry ABLE above, the content of <etym> consists of two <mentioned> elements, each with a <lang> and possibly a <gloss>, meaning it must be a root entry.

```
<entry xml: type="derivation">
  <form>...</form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.009999.graft">graft</mentioned>
    <pc>.</pc>
  </etym>
```

- 70 In the encoding of the entry APPLE-GRAFT, the content of <etym> consists of two <mentioned> elements, each with a @corresp attribute that points to other entries within the same dictionary, indicating a derivation. While the effort of identifying the target entry and inserting the corresponding @xml:id attribute is not insignificant, from our point of view the resulting network of linked entries is worth the effort.

4.6. Stepwise Refinement of <sense>: <num>, <def>, and <gramGrp> with <gram>

- 71 The function of <sense> as a container for the semasiological information of dictionary entries was explained the first half of this paper. Some sections of the encoding of ABLE can illustrate the flexibility of the concept of crystals for the encoding of complex semantic structures. The first step of refinement adds <num> elements to label the different <sense>s.

```
<entry xml: type="Root">
  <form> ... </form>
  <etym> ... </etym>
  <sense>
    <lb/><num>1.</num>
    <def>...</def><cit>...</cit><cit>...</cit>
  </sense>
  <sense>
    <lb/><num>2.</num>
    <def>Having power sufficient; enabled.</def>
    <cit type="example">
      <lb/><quote>All mankind acknowledge themselves able and
        sufficient to <lb/> do many things, which actually they never do.
      </quote>
      <bibl><author>South</author>'s <title>Serm.</title></bibl>
    </cit>
  </sense>
  <sense>
    <lb/><num>3.</num>
    <gramGrp>
      <gram type="syntax">Before a verb, with the participle
        <hi rend="italics">to</hi></gram>
    </gramGrp>,
    <def>it signifies generally hav-<lb/>ing the power</def>;
    <gramGrp>
      <gram type="syntax">before a noun, with <hi rend="italics">for</hi></gram>
```

```

    </gramGrp>,
    <def> it means <hi>qualified</hi></def>.
    <!-- instances of <cit type="example"> omitted for brevity -->
</sense>

```

- 72 In a second step—<num>3.</num>—one <sense> element is used to combine the morpho-syntactic features “*able + to* before a verb” in the <gramGrp> container with the semasiological definition “signifies generally having the power” contained in the <def> element. In a different construction with *able*, the morpho-syntactic feature “before a noun, with for” in <gramGrp> and <gram> is connected with the definition ‘it means qualified’ in <def>. Usually we find grammatical information in a kind of shorthand in the source, which is likewise encoded briefly:

```
<gramGrp><pos norm="noun">n.s.</pos></gramGrp>
```

- 73 For ABLE, we have a discursive example, which as such is interesting not only in its own right but also because it combines two clearly distinct syntactic structures and divergent semantic paraphrases into one sense. The <cit> examples that follow in sense number 3 repeat the structures and illustrate both usages:

```

<cit type="example">
  <lb/><quote>Wrath is cruel, and anger is outrageous; but who is
  able <lb/> to stand before envy?</quote>
  <bibl><title>Prov.</title>
    <biblScope type="part">xxvii.</biblScope>
    <biblScope type="ll">4.</biblScope>
  </bibl>
</cit>
<cit type="example">
  <lb/><quote>There have been some inventions also, which have been
  <lb/>able for the utterance of articulate sounds,
  as the speaking of <lb/>certain words.</quote>
  <bibl><author>Wilkin</author>'s <title>Mathematical Magic</title>.
  </bibl>
</cit>

```

- 74 The phrases *able to* and *able for* are marked by italics in the print dictionary, but this was not captured in the encoding. Furthermore, while the refinement of the encoding could be extended to word level and features of a fine-grain morpho-syntactical analysis, this is beyond what we want to illustrate in this paper. Therefore we have just encoded to support analysis of syntax.

4.7. Bilingual Dictionaries: A Shift of Perspective

- 75 The consistent modeling of heterogeneous lexical structures can be extended to the more complex structures we find in the two bilingual dictionaries, Adelung’s English-German translation of Johnson’s dictionary (1783–1796) and Ebers’ *New and Complete Dictionary of the German and English Languages* (1796), compiled using Adelung’s and Schwan’s lexicographical materials. Nevertheless a comparable precision in the encoding can be achieved. Let us first compare the entry *Apple-tart* in Johnson’s dictionary and Adelung’s translation:

```

<entry xml: type="derivation">
  <form type="lemma" norm="apple tart">
    <lb/><orth rend="smallcaps">APPLE-TART</orth><pc>.</pc>
    <gramGrp><pos norm="noun"/></gramGrp>
  </form>
  <etym><pc>[</pc>from
    <mentioned corresp="#sjdict1f.1755.000345.apple">apple</mentioned>
    <lbl>and</lbl>
    <mentioned corresp="#sjdict1f.1755.029999.tart">tart</mentioned>
    <pc>.]</pc>
  </etym>
  <sense>
    <def>A tart made of apples.</def>
  </sense>
</entry>

```

```

    <cit type="example">
      <lb/><quote>What, up and down carv'd like an apple-tart.</quote>
      <lb/><bibl><author>Shakespeare</author>'s
      <title>Taming of the Shrew</title>.
    </bibl>
  </cit>
</sense>
</entry>
<entry xml: type="main">
  <form xml:lang="en" type="lemma" norm="apple tart">
    <lb/><orth>'Apple-tart</orth><pc>,</pc>
    <gramGrp><pos norm="noun" xml:lang="la">subst.</pos></gramGrp>
  </form>
  <sense xml:lang="de">
    <def>eine Torte von Ä-<lb/>pfeeln,</def>
    <cit type="translation"><quote>eine Äpfeltorte.</quote></cit>
  </sense>
</entry>

```

76 In contrast to Johnson, Adelung, meeting the requirements of an English-German dictionary, left out the <etym> element on word-formation and the Shakespeare quotation and added the word-class information. He translated Johnson's definition of *apple-tart* almost literally into German and then added the slightly strange German compound *Äpfeltorte*.

77 The encoding of the translation becomes more complex because of the mix of two languages which requires an additional control of the extension and inheritance of the @xml:lang attribute. The use of the German plural form *Äpfel* in *Äpfeltorte* may have been inspired by Johnson's plural definition and the fact that a decent apple-tart requires more than one apple. Ten years later, in Adelung's monolingual German dictionary, the entry shows no umlaut and the definition is derived from a recipe that puts the sliced apples on top (1793–1801, vol. 1, 412).

78 In a final look at Ebers' German-English dictionary, the randomly chosen sample entry *fähig* shows the problems in encoding bilingual dictionaries when translation from mother-tongue into a foreign language is involved.

```

<entry xml: type="main">
  <form xml:lang="de" type="lemma" norm="fähig">
    <lb/><orth>Fähig</orth><pc>,</pc>
    <gramGrp><pos xml:lang="la" norm="adjective">adj. et adv.</pos>
  </gramGrp>
</form>
<sense>
  <def xml:lang="de">tüchtig, geschickt</def>
  <cit type="translation" xml:lang="en">
    <quote>capable, able, apt, fit, proper.</quote>
  </cit>
  <cit type="example" xml:lang="de">
    <quote>zu etwas fähig seyn,</quote></cit>
  <cit type="translation" xml:lang="en">
    <quote>to be capable or <lb/>fit for a Thing.</quote></cit>
  <lb/><cit type="example" xml:lang="de">
    <quote>sie ist des Erbrechts nicht fähig</quote></cit>
  <cit type="translation" xml:lang="en">
    <quote>she is <lb/>incapable for Succession.</quote></cit>
</sense>
<sense>
  <def xml:lang="de">fähig, lehrsam, gelehrig,</def>
  <cit type="translation" xml:lang="en">
    <quote>docile, teach- <lb/>able.</quote></cit>
  <lb/><cit type="example" xml:lang="de">
    <quote>fähig etwas zu erfinden</quote></cit>
  <cit type="translation" xml:lang="en">
    <quote>inventive.</quote></cit>
  <cit type="example" xml:lang="de">
    <quote>der Unterweisung fähig</quote></cit>
  <cit type="translation" xml:lang="en">
    <quote>susceptible of <lb/>Discipline, of Instruction</quote></cit>
  <lb/><cit type="example" xml:lang="de">

```

```

    <quote>er ist fähig alles zu unternehmen</quote></cit>
    <cit type="translation" xml:lang="en">
      quote>he <lb/>is a Man that will undertake any <lb/>Thing</quote></cit>
  </sense>
  <sense>
    <def xml:lang="de">fähig machen,</def>
    <cit type="translation" xml:lang="en">
      <quote>to enable or fit, to in- <lb/>capacitate, to habilitate.</quote>
    </cit>
    <lb/><cit type="example" xml:lang="de">
      <quote>der Hunger macht einen zu allem fähig,</quote></cit>
    <lb/><cit type="translation" xml:lang="en">
      <quote>Hunger breaks through Stone-<lb/>Walls, or Hunger drives
        the Wolf <lb/>out of the Forest.</quote></cit>
    <lb/><cit type="example" xml:lang="de">
      <quote>einen wieder fähig machen,</quote></cit>
    <cit type="translation" xml:lang="en">
      <quote>to rehabi-<lb/>litate, re-enable, re-instate, re- <lb/>store,
        or re-establish one</quote></cit>
  </sense>
</entry>

```

79 At first glance the main lexicographical problem here is to specify the different senses of *fähig*, first in German (with a separate `<sense>`, each containing a `<def>`, for each sense), then in translating the German adjectives into the English equivalents (using `<cit type="translation">`), and finally in adding English translations (in `<cit type="translation">`) of German example phrases (in `<cit type="example">`) containing the adjective. Unlike in Johnson's dictionary, the senses are not numbered and the principle of their order is not quite clear.

80 Recalling the longish title of Ebers' dictionary, *New and Complete Dictionary of the German and English Languages Composed Chiefly After the German Dictionaries of Mr. Adelung and of Mr. Schwan*, it is worthwhile taking a closer look at Ebers' possible sources. The entry *fähig* in Adelung's dictionaries (1774–1786, vol. 2; 1793–1801, vol. 2) is built around two numbered senses and looks completely different. But checking Christian Friedrich Schwan's *Nouveau dictionnaire de la langue allemande et française: Composé sur les dictionnaires de M. Adelung et de l'Acad. Française* (1782, 519) shows clearly how Ebers had compiled this entry of his dictionary:

```

<entry xml: type="main">
  <form xml:lang="de" rend="iso15924:Latn" type="lemma" norm="fähig">
    <lb/><orth>FÄHIG</orth><pc>,</pc>
    <pc></pc><orth rend="iso15924:Latf">fähig</orth><pc></pc>
    <gramGrp>
      <pos xml:lang="fr" norm="adjective">adj. & adv.</pos>
    </gramGrp>
  </form>
  <sense rend="iso15924:Latn">
    <def xml:lang="de">tüchtig, geschickt;</def>
    <cit type="translation" xml:lang="fr">
      <quote>Capable, habile, propre.</quote></cit>
    <cit type="example" xml:lang="de"><quote>Zu etwas fähig seyn;</quote></cit>
    <lb/><cit type="translation" xml:lang="fr">
      <quote>être capable de qq. ch. être propre à une chose.</quote></cit>
    <lb/><cit type="example" xml:lang="de">
      <quote>Sie ist des Erbrechts nicht fähig;</quote></cit>
    <cit type="translation" xml:lang="fr">
      <quote>elle n'est pas <lb/>habile à succéder.</quote></cit>
  </sense>
  <sense rend="iso15924:Latn">
    <abbr>It.</abbr><def xml:lang="de">Fähig, lehrsam, geleh-<lb/>rig</def>
    <cit type="translation" xml:lang="fr"><quote>docile.</quote></cit>
    <cit type="example" xml:lang="de">
      <quote>Fähig etwas zu erfinden;</quote></cit>
    <cit type="translation" xml:lang="fr"><quote>inven-<lb/>tif.</quote></cit>
    <cit type="example" xml:lang="de">
      <quote>Der Unterweisung fähig;</quote></cit>
    <cit type="translation" xml:lang="fr">

```

```

      <quote>susceptible de di-<lb/>scipline.</quote></cit>
<cit type="example" xml:lang="de">
  <quote>Er ist fähig alles zu unternèhmen;</quote></cit>
<lb/><cit type="translation" xml:lang="fr">
  <quote>il est homme à tout entreprendre.</quote></cit>
<cit type="example" xml:lang="de">
  <quote>Dinge, die<lb/>nicht jedermann zu verstehen fähig ist;</quote>
</cit>
<cit type="translation" xml:lang="fr">
  <quote>des <lb/>choses qui ne sont pas à la portée de tout
  le mon-<lb/>de</quote></cit>
<cit type="example" xml:lang="de">
  <quote>Er ist nicht fähig, euch in geringsten zu<lb/>schaden</quote></cit>
<cit type="translation" xml:lang="fr">
  <quote>il est incapable de vour nuire aucunement.</quote></cit>
<lb/><cit type="example" xml:lang="de"><quote>Fähig machen</quote></cit>
<cit type="translation" xml:lang="fr"><quote>habiliter.</quote></cit>
<cit type="example" xml:lang="de">
  <quote>Der Hunger macht <lb/>einen zu allem fähig;</quote></cit>
<cit type="translation" xml:lang="fr">
  <quote>la faim chasse le loup hors<lb/>du bois.</quote></cit>
<cit type="example" xml:lang="de">
  <quote>Einen wieder fähig machen;</quote></cit>
<cit type="translation" xml:lang="fr">
  <quote>rèhabi-<lb/>liter qq. un.</quote></cit>
</sense>
</entry>

```

- 81 With the exception of two phrases—“Dinge, die nicht jedermann zu verstehen fähig ist” and “Er ist nicht fähig euch in geringsten zu schaden”—Ebers has copied the German text of Schwan’s dictionary and replaced the French translation equivalents by English ones. The encoding problems remain the same and we think that the solution we propose is adequate.

5. Conclusion

- 82 Above we applied our encoding suggestions for the <form> block to Johnson’s entry *To APPLAUD* but did not comment on the unusual structure of the elements <sense> and <cit>: two numbered senses, followed by two quotations. A look at the last edition (the fourth folio edition of 1773), which was considerably revised and prepared for publication by Johnson himself, can make the author’s original intentions clearer. Thanks to Anne McDermott’s excellent CD-ROM edition, published in 1996, we have access to an SGML encoding of the texts of both the first and fourth folio editions and can not only compare the texts themselves but also the change over the years from TEI P3 SGML of 1994 to the current P5 using XML Schema:

- 83 **First folio edition [TEI P5]:**

```

<entry xml: type="Root" >
  <lb/><form type="lemma" norm="applaud">
    <gram type="infinitiveParticle">To</gram>
    <orth rend="allcaps">APPLA'UD</orth><pc>.</pc>
    <gramGrp><pos norm="verb">v.a.</pos></gramGrp>
  </form>
  <etym>
    <pc>[</pc><mentioned xml:lang="la">applaudo</mentioned><pc>,</pc>
    <lang><abbr>Lat.</abbr></lang><pc>]</pc>
  </etym>
  <lb/><sense>
    <num>1.</num>
    <def>To praise by clapping the hand.</def>
  </sense>
  <lb/><sense>
    <num>2.</num>
    <def>To praise in general.</def>
  </sense>
  <cit type="example">
    <lb/><quote>I would applaud thee to the very echo,
    <lb/>That should applaud again.</quote>

```

```

    <bibl><author><abbr>Shakesp.</abbr></author><title>Macbeth</title>.</bibl>
  </cit>
  <cit type="example">
    <lb/><quote>Nations unborn your mighty names shall sound,
    <lb/>And worlds applaud that must not yet be found!</quote>
    <bibl><author>Pope</author>.</bibl>
  </cit>
</entry>

```

84 Ann McDermott Fourth folio edition [TEI P3 SGML]:

```

<ENTRYFREE N="1999" TYPE="4">IV
  <FORM>
    <HI REND="ital">To</HI> <HI REND="acp">APPLA'UD.</HI>
  </FORM>
<PB SIG="Bb2r" MACFILE=":4:100:148.CAL" PCFILE="4100148.CAL">
  <POS><HI REND="ital">v.a.</HI></POS>
  <ETYM>[<HI REND="ital">applaudo,</HI> Lat.]</ETYM>
  <SENSE N="1">
    <DEF>
      <NUM>1.</NUM> To praise by clapping the hand.
    </DEF>
    <EG TYPE="verse">
      <QUOTE>
        <L>I would <HI REND="ital">applaud</HI> thee to the very echo,</L>
        <L>That should <HI REND="ital">applaud</HI> again.</L>
      </QUOTE>
      <AUTHOR><HI REND="ital">Shakesp.</HI></AUTHOR>
      <TITLE><HI REND="ital">Macbeth.</HI></TITLE>
    </EG>
  </SENSE>
  <SENSE N="2">
    <DEF>
      <NUM>2.</NUM> To praise in general.
    </DEF>
    <EG TYPE="verse">
      <QUOTE>
        <L>Nations unborn your mighty names shall sound,</L>
        <L>And worlds <HI REND="ital">applaud</HI> that must
          not yet be sound!</L>
      </QUOTE>
      <AUTHOR><HI REND="ital">Pope.</HI>
    </AUTHOR>
  </EG>
</SENSE>
</ENTRYFREE>

```

85 We can conclude:

1. The transcription of the entry APPLAUD in the SGML version of the fourth folio edition shows clearly that Johnson had intended to illustrate each definition with an illustrative quotation, as elsewhere in the dictionary, and that the unusual structure of the first folio text—two numbered senses, followed by two quotations—is simply a typesetting error.
2. Both encodings have many structural features in common: with the exception of <cit> and <pc>, all elements used in our encoding were available in TEI P3, whereas the mechanisms usable at the attribute level are not comparable. But the main difference is the style of the encoding: although the SGML version is very close to the typography of the text, our encoding, using crystals, aims more at interpreting typographical detail in order to capture lexicographic and linguistic data and to constrain encoding options in favor of robust interoperability and reusability of resources.

A. Appendix: Facsimiles

A 1 Johnson Entry "A B L"

A B L

must take the test; which is an *abjuration* of some doctrines of the church of Rome.

There is likewise another oath of *abjuration*, which laymen and clergymen are both obliged to take; and that is, to abjure the Pretender.

Wylliffe's Parergon Juris Canonici.
To ABLA'CTATE. *v. a.* [ablacto, Lat.] To wean from the breast.

ABLACTION. *n. f.* One of the methods of grafting; and, according to the signification of the word, as it were a weaning of a cyon by degrees from its mother stock, not cutting it off wholly from the stock, till it is firmly united to that on which it is grafted.

ABLAQUEA'TION. *n. f.* [ablaqueatio, Lat.] The act or practice of opening the ground about the roots of trees, to let the air and water operate upon them.

Trench the ground, and make it ready for the spring; prepare also soil, and use it where you have occasion: Dig borders. Uncover as yet roots of trees, where *ablaqueation* is requisite.

The tenure in chief ought to be kept alive and nourished; the which, as it is the very root that doth maintain this silver stem, that by many rich and fruitful branches spreadeth itself into the chancery, exchequer, and court of wards: so if it be suffered to starve, by want of *ablaqueation*, and other good husbandry, not only this yearly fruit will much decrease from time to time, but also the whole body and boughs of that precious tree itself, will fall into danger of decay and dying.

ABLA'TION. *n. f.* [ablatio, Lat.] The act of taking away.

ABLA'TIVE. *n. a.* [ablativus, Lat.]
1. That which takes away.
2. The sixth case of the Latin nouns; the case which, among other significations, includes the person from whom something is taken away. A term of grammar.

ABLE. *adj.* [habile, Fr. habilis, Lat. skilful, ready.]

1. Having strong faculties, or great strength or knowledge, riches, or any other power of mind, body, or fortune.

He was not afraid of an able man, as Lewis the Eleventh was. But, contrariwise, he was served by the *ablest* men that were to be found; without which his affairs could not have prospered as they did.

Such other gambol faculties he hath, that shew a weak mind and an *able* body, for the which the prince admits him: for the prince himself is such another: the weight of an hair will turn the scales.

2. Having power sufficient; enabled.

All mankind acknowledge themselves *able* and sufficient to do many things, which actually they never do.

Every man shall give as he is *able*, according to the blessing of the Lord thy God, which he hath given thee.

3. Before a verb, with the participle *to*, it signifies generally having the power; before a noun with *for*, it means *qualified*.

Wrath is cruel, and anger is outrageous; but who is *able* to stand before envy?

There have been some inventions also, which have been *able* for the utterance of articulate sounds, as the speaking of certain words.

To A'BLE. *v. a.* To make able; to enable, which is the word commonly used. See ENABLE.

Plate him with gold,
And the strong lance of justice hurtles breaks:
Arm it with rage, a pigmy's straw doth pierce it.
None does offend, none, I say none; I'll *able* 'em;
Take that of me, my friend, who have the power:
To seal th' accuser's lips.

ABLE-BODIED. *adj.* Strong of body.

It lies in the power of every fine woman, to secure at least half a dozen *able-bodied* men to his majesty's service.

To A'BLEGATE. *v. a.* [ablego, Lat.] To send abroad upon some employment; also to send a person out of the way that one is weary of.

ABLEGATION. *n. f.* [from *ablegate*.] A sending abroad, or out of the way.

ABLENES. *n. f.* [from *able*.] Ability of body or mind; vigour, force; faculties.

That nation doth so excel, both for comeliness and *ableness*, that from neighbour countries they ordinarily come, some to drive, some to learn, some to behold.

ABLEPSY. *n. f.* [Abληψία, Gr.] Want of sight, natural blindness; also unadvisedness.

ABLOU'ISH. *n. f.* [ablu'gash, Lat.] A prodigal spending on meat and drink.

To A'BLIGATE. *v. a.* [abligo, Lat.] To bind or tie up from.

To A'BLOCATE. *v. a.* [abluo, Lat.] To let out to hire.

Perhaps properly by him who has hired it from another.

ABLOCUTION. *n. f.* [from *ablocute*.] A letting out to hire.

To ABLU'ES. *v. a.* [abluo, Lat.] To be unlike.

ABLUNT. *adj.* [abluens, Lat. from *abluo*, to wash away.]

1. That which washes clean.

2. That which has the power of cleaning.

ABLU'TION. *n. f.* [ablutio, Lat.]

1. The act of cleansing, or washing clean.

A B O

There is a natural analogy between the *ablation* of the body and the purification of the soul; between eating the holy bread and drinking the sacred chalice, and a participation of the body and blood of Christ.

Wash'd by the briny wave, the pious train
Are cleans'd, and cast th' *ablutions* in the main.

2. The rinsing of chymical preparations in water, to dissolve and wash away any acrimonious particles.

3. The cup given, without consecration, to the laity in the popish churches.

To A'BNEGATE. *v. a.* [from *abnego*, Lat.] To deny.

ABNEGATION. *n. f.* [abnegatio, Lat. denial, from *abnego*, to deny.] Denial, renunciation.

The *abnegation* or renouncing of all his own holds and interests, and trusts of all that man is most apt to depend upon, that he may the more expeditely follow Christ.

ABNODA'TION. *n. f.* [abnodatio, Lat.] The act of cutting away knots from trees; a term of gardening.

ABNO'RMOUS. *adj.* [abnormis, Lat. out of rule.] Irregular, mishapen.

ABO'ARD. *adv.* [a sea term, but adopted into common language; derived immediately from the French *à bord*, as *aller à bord*, *envoyer à bord*. *Bord* is itself a word of very doubtful original, and perhaps, in its different acceptations, deducible from different roots. *Bopb*, in the ancient Saxon, signified a *house*; in which sense, to *go aboard*, is to take up residence in a ship.] In a ship.

Which, when far off, Cymocles heard and saw,
He loudly call'd to such as were *aboard*,
The little bark unto the shore to draw,

And him to ferry over that deep ford.

I made this answer, that he might land them, if it pleased him, or otherwise keep them *aboard*.

When morning rose, I sent my mates to bring
Supplies of water from a neighbouring spring;

Whilst I the motions of the winds explor'd;
Then summon'd in my crew, and went *aboard*.

ABO'DE. *n. f.* [from *abide*.]

1. Habitation, dwelling place of residence.

But I know thy *abode*, and thy going out, and thy coming in and thy rage against me.

Others may use the ocean as their road,
Only the English make it their *abode*;

Whose ready sails with every wind can fly,
And make a covenant with th' unconstant sky.

2. Stay, continuance in a place.

Sweet friends, your patience for my long *abode*;
Not I, but my affairs, have made you wait.

Making a short *abode* in Sicily the second time, landing in Italy, and making the war, may be reasonably judged the business but of ten months.

The woodcocks early visit, and *abode*
Of long continuance in our temperate climate,
Foretell a liberal harvest.

3. To make *abode*; to dwell, to reside, to inhabit.

Deep in a cave the Sibyl makes *abode*;
Thence full of fate returns, and of the god.

To ABO'DE. *v. a.* [See BODE.] To foretoken or foreshow; to be a prognostic, to be ominous. It is taken, with its derivatives, in the sense either of good or ill.

Every man,
After the hideous storm that follow'd, was
A thing inspir'd; and, not consulting, broke
Into a general prophecy, that this tempest,
Destiny the garment of this peace, *aboded*
The sudden breach of it.

ABO'DEMENT. *n. f.* [from *to abide*.] A secret anticipation of something future; an impression upon the mind of some event to come; prognostication; omen.

I like not this.
For many men that stumble at the threshold,
Are well foretold that danger lurks within.—
—Tush! man, *abodements* must not now affright us.

My lord bishop, being somewhat troubled, took the freedom to ask him, Whether he had never any secret *abodement* in his mind? No, replied the duke: but I think some adventure may kill me as well as another man.

To ABO'LISH. *v. a.* [from *aboleo*, Lat. to blot out.]

1. To annul.

For us to *abolish* what he hath established, were presumption most intolerable.

On the parliament's part it was proposed, that all the bishops, deans, and chapters, might be immediately taken away, and *abolished*.

2. To put an end to; to destroy.

The long continued wars, between the English and the Scots, had then raised invincible jealousies and hate, which long continued peace hath since *abolished*.

Facsimile A.1.: Page with entry "ABLE" from Johnson (1755).

A.2. Johnson, Entries "To APPLAUD" and "APPLE"

A P P

3. Violent longing; eagerness after any thing.
No man could enjoy his life, his wife, or goods, if a mightier man had an *appetite* to take the fame from him. *Davies on Ireland*.
Hopton had an extraordinary *appetite* to engage Waller in a battle. *Clarendon*, b. viii.
Power being the natural *appetite* of princes, a limited monarch cannot gratify it. *Swift*.
4. Keenness of stomach; hunger; desire of food.
There be four principal causes of *appetite*; the refrigeration of the stomach, joined with some dryness; contraction; vellefaction, and absterfion; besides hunger, which is an emptiness. *Bacon's Natural History*, No 831.
5. It has sometimes of before the object of desire.
The new officer's nature needed some restraint to his immoderate *appetite* of power. *Clarendon*.
6. Sometimes to.

We have generally such an *appetite* to praise, that we greedily suck it in. *Government of the Tongue*, § 8.

APPETITION. *n. f.* [*appetitus*, Lat.] Desire.
The actual *appetition* or fastening our affections on him.

APPETITIVE. *adj.* [from *appetite*.] That which desires; that which has the quality of desiring.
The will is not a bare *appetitive* power as that of the sensual appetite, but is a rational appetite. *Hale's Origin of Mankind*.

TO APPLAUD. *v. a.* [*applaudo*, Lat.]

1. To praise by clapping the hand.
2. To praise in general.

I would *applaud* thee to the very echo,
That should *applaud* again. *Shakespeare, Macbeth*.
Nations unborn your mighty names shall found,
And worlds *applaud* that must not yet be found! *Pope*.

APPLAUDER. *n. f.* [from *applaud*.] He that praises or commends.

I had the voice of my single reason against it, drowned in the noise of a multitude of *applauders*. *Glanville's Scylla*.

APPLAUSE. *n. f.* [*applausus*, Lat.] Approbation loudly expressed; praise.

This general *applause*, and cheerful shout,
Argues your wisdom and your love to Richard. *Shak. R. III*.

Sylla wept,
And chid her barking waves into attention;
And fell Charybdis murmur'd soft *applause*. *Milton's Comus*.

Those that are so fond of *applause*, how little do they taste it when they have it? *South*.

See their wide streaming wounds; they neither came
For pride of empire, nor desire of fame;
Kings fight for kingdoms, madmen for *applause*,

But love for love alone, that crowns the lover's cause.

Dryden's Fables.

APPLE. *n. f.* [æppel, Saxon.]

1. The fruit of the apple tree.
Tall thriving trees confess'd the fruitful mold;
The redning *apple* ripens here to gold. *Pope's Odyssey*.
2. The pupil of the eye.

He instructed him; he kept him as the *apple* of his eye.
Deut. xxxii. 10.

APPLE OF LOVE.

Apples of love are of three sorts; the most common having long trailing branches, with rough leaves and yellow joints, succeeded by apples, as they are called, at the joints, not round, but bunched; of a pale orange shining pulp, and seeds within.

Mortimer's Art of Husbandry.

APPLE-GRAFT. *n. f.* [from *apple* and *graft*.] A twig of apple tree grafted upon the stock of another tree.

We have seen three and twenty sorts of *apple-grafts* upon the same old plant, most of them adorned with fruit. *Boyle*.

APPLE-TART. [from *apple* and *tart*.] A tart made of apples.

What, up and down carv'd like an *apple-tart*.
Shakespeare's Taming of the Shrew.

APPLE TREE. *n. f.* [from *apple* and *tree*.]

The fruit of this tree is for the most part hollowed about the foot stalk; the cells inclosing the seed are separated by cartilaginous partitions; the juice of the fruit is sowerish, the tree large and spreading; the flowers consist of five leaves, expanding in form of a rose. There is a great variety of these fruits.

Those for the dessert are, the white juneting, Margaret apple, summer pearmain, summer queening, embroidered apple, golden reinette, summer white Colville, summer red Colville, silver pippin, aromatick pippin, the gray reinette, la haute-bonté, royal russeting, Wheeler's russet, Sharp's russet, spice apple, golden pippin, nonpareil, and l'api. Those for the kitchen use are, codling, summer marigold, summer red pearmain, Holland pippin, Kentish pippin, the hanging body, Loan's pearmain, French reinette, French pippin, royal russet, monstrous reinette, winter pearmain, pomme violette, Spencer's pippin, stone pippin, oakenpippin. And those generally used for cyder are, Devonshire royal wilding, redstreaked apple, the whitfour, Herefordshire underleaf, John apple, &c.

Thus *apple trees*, whose trunks are strong to bear
Their spreading boughs exert themselves in air. *Dryden*.

APPLE WOMAN. *n. f.* [from *apple* and *woman*.] A woman that sells apples.

A P P

Yonder are two *apple women* scolding, and just ready to uncoif one another. *Arbutnot and Pope's Mart. Scribl.*

APPLIABLE. *adj.* [from *apply*.] That which may be applied. For this word the moderns use *applicable*; which see.

Limitations all such principles have, in regard of the varieties of the matter whereunto they are *applicable*. *Hooker*, b. v.

All that I have said of the heathen idolatry is *applicable* to the idolatry of another sort of men in the world. *South*.

APPLIANCE. *n. f.* [from *apply*.] The act of applying; the thing applied to.

Diseases despo'rate grown,
By desperate *appliances* are relieved. *Shakespeare, Hamlet*.

APPLICABILITY. *n. f.* [from *applicable*.] The quality of being fit to be applied to something.

The action of cold is composed of two parts; the one pressing, the other penetration, which require *applicability*. *Digby*.

APPLICABLE. *adj.* [from *apply*.] That which may be applied, as properly relating to something.

What he says of the portrait of any particular person, is *applicable* to poetry. In the character, there is a better or a worse likeness; the better is a panegyrick, and the worse a libel.

Dryden's Dufresnoy, Preface.

It were happy for us, if this complaint were *applicable* only to the heathen world. *Rogers*.

APPLICABLENESS. *n. f.* [from *applicable*.] Fitness to be applied.

The knowledge of facts may possibly, by that little part which we have already delivered of its *applicableness*, be of use in natural philosophy. *Boyle*.

APPLICABLY. *adv.* [from *applicable*.] In such a manner as that it may be properly applied.

APPLICATE. *n. f.* [from *apply*.] A right line drawn across a curve, so as to bisect the diameter thereof. *Chambers*.

APPLICATION. *n. f.* [from *apply*.]

1. The act of applying any thing to another; as, he mitigated his pain by the *application* of emollients.
2. The thing applied; as, he invented a new *application*, by which blood might be staunch'd.
3. The act of applying to any person, as a solicitor, or petitioner.

It should seem very extraordinary, that a patent should be pass'd, upon the *application* of a poor, private, obscure mechanic. *Swift*.

4. The employment of any means for a certain end.

If a right course be taken with children, there will not be much need of the *application* of the common rewards and punishments. *Locke*.

5. Intenfeness of thought; close study.

I have discovered no other way to keep our thoughts close to their business, but by frequent attention and *application*, getting the habit of attention and *application*. *Locke*.

6. Attention to some particular affair; with the particle *to*.

His continued *application* to such publick affairs, as may conduce to the benefit of his kingdom, diverts him from pleasures. *Addison's Freeholder*, No 46.

This crime certainly deserves the utmost *application* and punishment of a people to prevent it. *Addison*.

7. The condition of being used as means to an end.

There is no flint which can be set to the value or merit of the sacrificed body of Christ; it hath no measured certainty of limits, bounds of efficacy unto life it knoweth none, but is also itself infinite in possibility of *application*. *Hooker*, b. v.

This principle acts with the greatest force in the worst *application*; and the familiarity of wicked men more successfully debauches, than that of good men reforms. *Rogers*.

APPLICATIVE. *adj.* [from *apply*.] That which applies.

The directive command for counsel is in the understanding, and the *applicative* command for putting in execution, is in the will. *Bramhal against Hobbes*.

APPLICATORY. *adj.* [from *apply*.] That which comprehends the act of application.

APPLICATORY. *n. f.* That which applies.

There are but two ways of applying the death of Christ: faith is the inward *applicatory*, and if there be any outward, it must be the sacraments. *Taylor's Worshy Communicant*.

TO APPLY. *v. a.* [*aplico*, Lat.]

1. To put one thing to another.
2. To lay medicaments upon a wound.

Apply some speedy cure, prevent our fate,
And succour nature ere it be too late. *Addison's Ovid's Met*.

God has address'd every passion of our nature, *applied* remedies to every weakness, warn'd us of every enemy. *Rogers*.

3. To make use of as relative or suitable to something.

This brought the death of your father into remembrance, and I repeated the verses which I formerly *applied* to him. *Dryden's Fables, Ded*.

4. To put to a certain use.

The profits thereof might be applied towards the support of the year. *Clarendon*.

5. To use as means to an end.

These glorious beings are instruments in the hands of God, who

Facsimile A.2.: Page with entries "To APPLAUD" and "APPLE" from Johnson (1755).

A.3. Adelung, Entry "Apple"



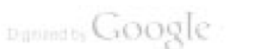
Facsimile A.3.: Page with entry "Apple" from Adelung (1783-1796).

A.4. Ebers, Entry "FÄHIG"

Fad	(824)	Fah
das Holz zu Faden schlagen, to cord, to fathom Wood; to measure it by the Cord.		Fagott, der, the Balloon.
der Faden, Rahmen, die Klasten abzu-messen, a Cord to measure Wood.		Fähe, die, (Term of the Hunters) the Female of Dogs and of all quadrupede Beasts of Prey.
ein Faden Rahmholz, a Cord of Wood.		Fahngulden, der, see Bildfang.
Fadenbrüche, die, (Term of the Weavers) slipping, gliding or running Thread.		Fahen, verb. irreg. ad. (better fangen) to take, to catch; see Fangen.
Fadenflege, die, the Scatopse.		Fähig, adj. et adv. tüchtig, geschickt, capable, able, apt, fit, proper.
Fadenhaken, der, (in Silk Manufactures) a bent Wire on Silk-Reels.		zu etwas fähig seyn, to be capable or fit for a Thing.
Fadenholz, das, see Klastenholz; Wood that is measured and sold by the Cord.		sie ist des Erbrechts nicht fähig, she is incapable for Succession.
Fadenfaser, der, the Hispa Antennae pelues.		fähig, lehrsam, gelehrtig, docile, teachable.
Fadenfraut, das, the Cotton Herb.		fähig etwas zu erfinden, inventive.
Fadenleiter, der, a perpendicular Piece of a Silk-Reel, to lead the two Threads on the Reel.		der Unterweisung fähig, susceptible of Discipline, of Instruction.
Fadennackend, adj. et adv. as bare as the Hand, naked as a Worm, without a Cover, threadbare.		er ist fähig alles zu unternehmen, he is a Man that will undertake any Thing.
Fadennudeln, die, Vermicelli.		fähig machen, to enable or fit, to incapacitate, to habitate.
Fadenrecht, adj. et adv. dem Faden gemäß, dem Faden nach, as even as a Thread, straight forward, in a straight Line with the Thread; perpendicular.		der Hunger macht einen zu allem fähig, Hunger broaks through Stone-Walls, or Hunger drives the Wolf out of the Forest.
fadenrecht herunter hangen, to hang downright, perpendicularly.		einen wieder fähig machen, to rehabilitate, re-enable, re-instate, restore, or re-establish one.
die Leinwand nach dem Faden schneiden, to cut the Cloth or Linnen even with the Thread, or as even or straight as the Thread runs.		Fähigkeit, die, Ability, Capacity, Skill, Sufficiency, Docility, Genius, Talent, Disposition for something, Faculty, Comprehension, Conception, Discernment, Intelligence, Reach.
gegen den Faden schneiden, to cut a slope, across or overthwart.		dieser Knabe hat eine große Fähigkeit, this Boy has a great Capacity, a good Genius.
Fadenscheinig, adj. et adv. was den Faden durchscheinen läßt, that shews the Thread.		ein Mensch von geringer Fähigkeit, a Man of mean Parts, of a shallow Brain.
Fadenstüber, das, Silver-Wire; a Silk-Thread spun over with Silver.		dieser Mensch hat in seiner Profession große Fähigkeit, this Man is very well skilled or has great Skill in his Profession or in his Art.
Fadenwurm, der, a long thin Worm like or resembling a Thread; (Gordius aquaticus Linn).		so weit erstreckt sich seine Fähigkeit nicht, that's above his Reach or Capacity.
Fadengucca, die, the Virginian filaceous Yucca; (Yucca filamentosa Linn).		Fahl, adj. et adv. fahl, bleich, blaß, schwärzlich-grau, fallow, fallow yellowish, pale, reddish or between red and yellow, tawny.
Fadig or Fadiht, adj. et adv. filaceous, that has Filaments.		ein fahles Pferd mit pferchblüthfarbenen Haaren, oder weiß mit braunen Flecken, a flea-bitten gray, or a dapple gray Horse; a white Horse spotted with Sorrel and Bay Spots.
Fädig, adj. et adv. gleichfädeniges, ungleichfädeniges Tuch, Cloth or Linnen that has an even, an uneven Thread.		Fahlerz, das, a Silver Ore of a gray Colour mixed with brown.
dreifädenige oder dreydrätige Strümpfe, three threaded Stockings, or Stockings of three Threads.		Fählich,

A.5. Schwan, Entry "FÆHIG"

FAD.	FAH.	519
<p>de commission. Eine Factorei-Handlung; une <i>factorerie, factorie</i>.</p> <p>FACTUR, (factur) (die) gen. der-, pl. die-en; <i>La facture</i>. Das Facturbuch; le <i>mémoire des marchandises</i>.</p> <p>FACULTÆT, (faculté) (die) gen. der-, pl. die-en; <i>La faculté</i>. Die theologische, juristische, medicinische und philosophische Facultät; la <i>faculté de théologie, de droit, de médecine, des arts</i>. Der Facultist; un <i>membre de la faculté</i>.</p> <p>FÄDELN, (fädeln) vey. EINFÄDELN.</p> <p>FADEN, (faden) (der) gen. des-s, pl. die-n; <i>Le fil</i>. Ein dünner Faden; un <i>fil</i>. Ein Faden von Wolle, Seide; un <i>fil de laine, de soie</i>. Bindfaden; <i>de la ficelle</i>. Faden im Tuche, Zuvorge; <i>la suture, corde, le fil</i>. Man kennt das Tuch am Faden; on <i>connaît le drap à la suture</i>. Wenn das Tuch abgetragen ist, so sieht man den Faden; <i>quand le drap est usé, il montre la corde</i>. Zu Fäden schlagen, mit weitläufigen Stichen verloren zusammenheften; <i>jauster</i>, Faden womit man flickt; <i>la carisane</i>. Erster Faden von gefalteter Seide; <i>la centaine</i>. En <i>T. de Boton</i>, on <i>appelle</i> Faden; <i>les filaments</i>. Unsern lieben Frauen Faden; <i>le fil de la vierge, les arandelles, filandres, certains fils blancs</i> &c. lange qui <i>volent en l'air dans les beaux jours d'Automne</i>. Der Faden in der Fahre des Hirsches; <i>la combette, fente qui est au milieu du pied du cerf</i>. Der Faden an einem Messer; <i>le morsil</i>; <i>certaines petites parties d'acier presque imperceptibles, qui restent au tranchant d'un couteau, d'un rasoir</i> &c. lorsqu'on les a <i>passés sur la meule</i>. Einem Schärmeßer den Faden nehmen; <i>ôter le morsil d'un rasoir</i>.</p> <p>On <i>dit figur.</i> Nicht einen trocknen Faden am Leibe haben; <i>être tout mouillé, perir jusqu'à la peau</i>. Er hat nicht einen Faden weiß Zeug; <i>il n'a point de linge blanc</i>. Der Lebensfaden; <i>la vie, le cours de la vie, le fil de la vie, les jours d'une personne</i>. Den Lebensfaden abreißen; <i>donner la mort, abréger la vie</i>. Sein Leben hängt an einem Faden; <i>il est en grand danger de la vie</i>.</p> <p>Der Faden, Zusammenhang einer Geschichte, Rinde; <i>le fil, la suite d'un récit, d'une histoire, d'un discours</i>. <i>fi.</i> Der Faden; <i>la brasse mesure de la longueur de deux bras tendus</i>. Der Hafen ist drei Faden tief; <i>le port a trois brasses d'eau</i>. Man legte sich auf sechs Faden Wassers vor Anker; <i>on mouilla sur six brasses</i>. <i>fi.</i> Faden; (<i>T. forell.</i>) Klistert; <i>la toise, corde</i>. Zehn Faden lang; <i>long de dix toises</i>. Das Holz zu Fäden schlagen; <i>corder le bois</i>. Das krumme Holz laßt sich nicht gut in Faden setzen; <i>le bois tortu ne se corde pas bien</i>. Der Faden, Rahmen, die Klistert abzumessen; <i>la membrure</i>. Ein Faden Rahmholz; <i>une membrure de bois</i>.</p>	<p>Die Fadenbrüche; (<i>T. de Tisser.</i>) <i>les fils courants, braves courantes</i>. Die Fadenlinge; <i>la fatopé, épée de mouche</i>. Der Fadenhalter; <i>l'arquet, les arquets</i>. Der Fadenkäser; <i>le kips d'automne pelues</i>. Das Fadenkraut; <i>l'herbe à coton</i>. Der Fadenleiter; <i>le guide du dévidoir</i>. Fadennackend; <i>adj. & adv. nu comme la main, nu comme un cer</i>; <i>adv. à nu, sans voile, à découvert</i>. Fadennudeln; <i>des vermicelles</i>. Faderecht, <i>adj. & adv. dem Faden gemäß, dem Faden nach; de droit fil, perpendiculaire, perpendiculairement, à plomb</i>. Die Leinwand nach dem Faden schneiden; <i>couper la toile de (à) droit fil</i>. Gegen den Faden; <i>couper de biais</i>. Fadenscheinig, <i>adj. & adv. was den Faden durchscheinen läßt; qui montre la corde</i>. Fadenüber; <i>de l'argent trait, argent filé</i>. Der Fadenwurm; <i>le goudius, ver menu comme un fil, le dragonneau, dracoste</i>. Die Fadenyucca; <i>le yucca filamenteux de Virginie</i>. Fadicht; <i>adj. & adv. filamenteux</i>.</p> <p>FÄDENIG, (fädig) <i>adj. & adv. gleichfädiges, ungleichfädiges Tuch; Du drap, de la toile qui a le fil égal-égal</i>. Dreifädig; <i>ou dreidritige Strümpfe; des bas à trois fils</i>.</p> <p>FAGOTT, (fagott) (der) gen. des-es, pl. die-e; <i>Le bajon, une basse de haut-boys</i>.</p> <p>FÄHE, (fähe) (die) gen. der-, pl. die-n; (<i>T. de chasse</i>) <i>La femelle</i>.</p> <p>FAHEN, (fahen) v. a. irr. Ich fähe, du fähst, er fähst, wir fähen &c. <i>Imp.</i> Ich fieng, (subj.) fieng, i. h. gefahen ou gefangen; <i>impér.</i> fang; (<i>peu us.</i>) On <i>dit ordinairement</i> fangen; <i>Prendre; attraper</i>. Ein Falke der gut fähst, <i>mieux fangt; l'oiseau bon voleur</i>.</p> <p>FÆHIG, (fähig) <i>adj. & adv. tüchtig, geschickt; Capable, habile, propre</i>. Zu etwas fähig seyn; <i>être capable de qq. ch. être propre à une chose</i>. Sie ist des Erbrechts nicht fähig; <i>elle n'est pas habile à succéder</i>. <i>fi.</i> Fähig, lehrsam, gelehrt; <i>docile</i>. Fähig etwas zu erfinden; <i>inventif</i>. Der Unterweisung fähig; <i>susceptible de discipline</i>. Er ist fähig alles zu unternehmen; <i>il est homme à tout entreprendre</i>. Dinge, die nicht jedermann zu verstehen fähig ist; <i>des choses qui ne sont pas à la portée de tout le monde</i>. Er ist nicht fähig, auch in geringsten zu schaden; <i>il est incapable de vous nuire aucunement</i>. Fähig machen; <i>habiliten</i>. Der Hunger macht einen zu allem fähig; <i>la faim change le loup hors du bois</i>. Einen wieder fähig machen; <i>réhabiliten qq. un</i>.</p> <p>FÆHIGKEIT, (fähigkeit) (die) gen. der-, pl. die-en; <i>La capacité, habileté, docilité, le génie, talent; la disposition à qq. ch. faculté, compréhension, conception; it. discernement, l'intelligence, la portée</i>. Dieser Krabe hat eine große Fähigkeit; <i>ce gargon a une grande capacité</i>.</p>	



Facsimile A.5.: Page with entry "FÆHIG" from Schwan (1782).

Bibliography

Adams, V. 1976. *An Introduction to Modern English Word-Formation*. London: Longman.

Adelung, J. C. 1774–1786. *Versuch eines vollständigen grammatisch-kritischen Wörterbuches Der Hochdeutschen Mundart, mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen*. 5 vols. Leipzig: Breitkopf.

Adelung, J. C. 1783–1796. *Neues grammatisch-kritisches Wörterbuch der Englischen Sprache für die Deutschen; vornehmlich aus dem größern englischen Werke des Hrn. Samuel Johnson nach dessen vierten Ausgabe gezogen und mit vielen Wörtern, Bedeutungen und Beyspielen vermehrt*. 2 vols. Leipzig: im Schwickertschen Verlage.

Adelung, J. C. 1793–1801. *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart, mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen*, von Johann Christoph Adelung, Churfürstl. Sächs. Hofrath und Ober-Bibliothekar . . . 4 vols. Leipzig: Breitkopf.

- Atkins, S., N. Bel, F. Bertagna, P. Bouillon, N. Calzolari, C. Fellbaum, R. Grishman, R. Lenci, C. MacLeod, M. Palmer, G. Thurmaier, M. Villegas, and A. Zampolli. 2002. "From Resources to Applications. Designing the Multilingual ISLE Lexical Entry." In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, 687–693.
- Ebers, J. 1796. *New and Complete Dictionary of the German and English Languages composed chiefly after the German Dictionaries of Mr. Adelung and of Mr. Schwan*. . . . Vol. 1. Leipzig: Breitkopf and Haertel.
- Halpern, J. 2006. "The role of lexical resources in CJK natural language processing." In *Proceedings of the Workshop on Multilingual Language Resources and Interoperability*, 9–16.
- Henne, H., ed. 2001. *Deutsche Wörterbücher des 17. und 18. Jahrhunderts. Einführung und Bibliographie*. Hildesheim/Zürich/New York: Olms.
- Ide, N., A. Kilgarriff, and L. Romary. 2000. "A Formal Model of Dictionary Structure and Content." In *Proceedings of Euralex 2000*. Stuttgart, 113–126. <http://hal.archives-ouvertes.fr/hal-00164625>.
- Johnson, S. 1755. *A Dictionary of the English Language*. . . . 2 vols. London: W. Strahan.
- Kilgarriff, A. and D. Tugwell. "Sketching Words." *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins*, ed. Marie-Hélène Corréard. Stuttgart: EURALEX. 125–137. <http://www.kilgarriff.co.uk/Publications/2002-KilgTugwell-AtkinsFest.pdf>.
- Landau, S. I. 2001. *Dictionaries. The Art and Craft of Lexicography*. 2nd ed. Cambridge: Cambridge University Press.
- Lewis, D. 2012. *Die Wörterbücher von Johannes Ebers. Studien zur frühen englisch-deutschen Lexikographie*. PhD diss., University of Würzburg (in print).
- Luna, P. 2005. "The typographic design of Johnson's Dictionary." In *Anniversary Essays on Johnson's Dictionary*, ed. Jack Lynch and Anne McDermott, 175–197. Cambridge: Cambridge University Press.
- McDermott, A. ed. 1996. *Samuel Johnson, A Dictionary of the English Language*, on CD-ROM. The First and Fourth Editions. Cambridge: Cambridge University Press.
- Miller George A. and Christiane Fellbaum. 2007. "WordNet Then and Now." In *Language Resources and Evaluation* 41: 209–214, doi:10.1007/s10579-007-9044-6.
- Osselton, N. E. 2005. "Hyphenated Compounds in Johnson's Dictionary." *Anniversary Essays on Johnson's Dictionary*, eds. Jack Lynch and Anne McDermott, 160–174. Cambridge: Cambridge University Press.
- Reddick, A. 2006. *The Making of Johnson's Dictionary 1746–1773*. Rev. ed. Cambridge: Cambridge University Press.
- Romary, L. 2009. "ODD as a generic specification platform." Paper presented at *Text Encoding in the Era of Mass Digitization: Conference and Members' Meeting of the TEI Consortium*. <http://hal.inria.fr/inria-00433433>.
- Schwan, C. F. 1782. *Nouveau Dictionnaire de la Langue Allemande et Française*. . . . Vol. 1. Mannheim: Chez C.F. Schwan et M. Fontaine.
- TEI Consortium. 2012. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.0.2. Last updated February 2. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.

Notes

- 1 Some LMF packages, such as the description of subcategorization frames, do not yet have any equivalence in the TEI vocabulary, but the TEI extension mechanisms do facilitate the description of such extensions.
- 2 Note that some of the changes proposed in this paper (in particular regarding the systematic use of <sense>) have already been integrated into the December 2011 release (2.0.0, *Laurentian*).
- 3 Ideally, this should correspond to `model.formPart`, but in the current version of the TEI Guidelines this class is cluttered with other components which are there for purely syntactic (practical) reasons. We would limit this class to form <orth>, <pron>, <hyph>, <syll>, and <stress>.
- 4 Even if this is not allowed in the <entry> element, form representations still appear in: <cit>, <dictScrap>, <entryFree>, and <nym>, because of their membership to `model.entryPart`.

5 <http://www.isocat.org/>

6 Paul Luna's analyses here the typography of Johnson's folio edition of his dictionary (in opposition to different typography and text structure in the quarto and octavo editions). Folio is the old measure of size of a book and an indispensable term for research on Johnson's dictionaries.

7 Since Adelung's name does not appear on the title page nor elsewhere in the front matter, his role as a translator is little known. It is worth mentioning the publication context. Adelung studied and translated Johnson's dictionary while working on the two editions of his own German dictionaries. The first volume of his translation, containing the letters A to J, was published in 1783. This was after nearly three years of work—according to his preface (p. xii)—and before he finished the fifth and last volume of the first edition of his German dictionary which he had started in 1773 (Adelung 1774–1786). Thirteen years later, in 1796, he published the second volume of his translation with the letters K to Z, after having finished the first two volumes of the second and final edition of his German dictionary (Adelung 1793–1801). Against this background, future research into structural relations between Johnson's *Dictionary of the English Language* and Adelung's German dictionaries looks promising.

8 We do not encode the two typefaces for Latin script used by German printers of Adelung's and Ebers' dictionaries because there is a fixed relation between language (encoded using @xml:lang) and the typeface: for German texts the Fraktur variant is used, whereas for other languages Antiqua is used. We only encode exceptions to this rule, such as in Schwan's German-French dictionary, where ISO 15924 codes are used for the representation of names of scripts. We do not encode the indentation and alignment structure, nor do we encode italics in the contexts of part-of-speech labels (in a <pos> element), of cited forms in <etym> (if printed in italics), of the lemmata used in illustrative quotations (in a <cit> element), or of the names of authors and their works in the sources for the illustrative quotations (in a <bibl> element).

Cite this article

Electronic reference

Laurent Romary and Werner Wegstein, « Consistent Modeling of Heterogeneous Lexical Structures », *Journal of the Text Encoding Initiative* [Online], Issue 3 | November 2012, Online since 15 October 2012, connection on 05 November 2012. URL : <http://jtei.revues.org/540> ; DOI : 10.4000/jtei.540

Authors

Laurent Romary

Laurent Romary is *Directeur de Recherche* for INRIA (France) and guest scientist at Humboldt University (Berlin, Germany). He carries out research on the modeling of semi-structured documents, with a specific emphasis on texts and linguistic resources. He received a PhD degree in computational linguistics in 1989 and his *Habilitation* in 1999. During several years he launched and directed the *Langue et Dialogue* team at Loria (Nancy, France) and participated in several national and international projects related to the representation and dissemination of language resources and on man-machine interaction, coordinating the MLIS/DHYDRO, IST/MIAMM, and eContent/Lyrics projects. He has been the editor of ISO standard 16642 (TMF – Terminological Markup Framework) and is the chairman of ISO committee TC 37/SC 4 on Language Resource Management, as well as member (2001–2007) then chair (2008–2011) of the TEI Council. In the recent years, he led the Scientific Information directorate at CNRS (2005–2006) and established the Max-Planck Digital Library (Sept. 2006–Dec. 2008). He currently contributes to the establishment and coordination of the DARIAH infrastructure in Europe as transitional director.

Werner Wegstein

Werner Wegstein is a professor emeritus of German Linguistics and Computational Philology. His publications include a scholarly edition of an Old-High German Glossary (Ph.D. 1985), the first complete reverse index to a Middle High German dictionary (1990, together with E. Koller and N.R. Wolf), a *Habilitation* on computer-based philology (1995), conference papers on the application of IT to medieval German (2001), and a recently co-authored work on corpus linguistics (*Korpuslinguistik deutsch: synchron, diachron, kontrastiv*, 2005). He hosted the TEI workshops in Würzburg; is a founding member of TextGrid, the humanities partners of the German D-Grid initiative; and at present

is active in a project researching the interaction between the sciences and humanities in the field of variation.

Copyright

TEI Consortium 2012 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

Abstract

Our paper outlines a proposal for the consistent modeling of heterogeneous lexical structures in semasiological dictionaries, based on the element structures described in detail in chapter 9 (Dictionaries) of the TEI Guidelines. The core of our proposal describes a system of relatively autonomous lexical “crystals” that can, within the constraints of the relevant element’s definition, be combined to form complex structures for the description of morphological form, grammatical information, etymology, word-formation, and meaning for a lexical structure. The encoding structures we suggest guarantee sustainability and support re-usability and interoperability of data. This paper presents case studies of encoding dictionary entries in order to illustrate our concepts and test their usability. We comment on encoding issues involving <entry>, <form>, <etym>, and on refinements to the internal content of <sense>.

Index terms

Keywords : Dictionary encoding, semasiological dictionary, entry, form, sense, Samuel Johnson, Dictionary of the English Language

Author's notes

The authors would like to thank the reviewers of earlier versions of this paper, especially reviewer A, for their very detailed analysis and constructive criticism that contributed to the profile of our paper.